Low-Frequency Robust Cointegration Testing[*]

Ulrich K. Müller

Princeton University

Department of Economics

Princeton, NJ, 08544

Mark W. Watson

Princeton University

Department of Economics

and Woodrow Wilson School

Princeton, NJ, 08544

June 2007

(This version: April 2008)

Abstract

Standard inference in cointegrating models is fragile for two distinct reasons. First, inference assumes an $I(1)$ model for the common stochastic trends, which may not accurately describe the data's persistence; second, while cointegration concerns low-frequency variability, inference relies on higher frequency variability in the data. This paper discusses efficient inference about cointegrating vectors that is robust to both sources of misspecification. A simple test motivated by the analysis in Wright (2000) is developed and shown to be approximately optimal in the case of a single cointegrating vector.

JEL classification: C32, C12

Keywords: least favorable distribution, power envelope, nuisance parameters

1 Introduction

The fundamental insight of cointegration is that while economic time series may be individually highly persistent, some linear combinations are much less persistent. Accordingly, a suite of practical methods have been developed for conducting inference about cointegrating vectors, the coefficients that lead to this reduction in persistence. In their standard form, these methods assume that the persistence is the result of common $I(1)$ stochastic trends, and that the error correction terms, the non-persistent linear combination of the variables, are $I(0)$.[1] This $I(1)/I(0)$ dichotomy drives standard cointegration analysis, but may lead to fragile inference for two distinct reasons. First, the persistence reduction associated with moving from an $I(1)$ to an $I(0)$ process might be implausible in many applications. Second, standard methods rely critically on particular properties of the $I(1)$ process about which there may be uncertainty that cannot be resolved by examination of the data. This paper studies efficient inference methods for cointegrating vectors that are robust to both of these potential fragilities.

Consider first the issue that in the standard asymptotic reasoning, the error correction term and the stochastic trend are of different orders of persistence: apart from a scaling factor, the asymptotic behavior of $I(0)$ processes is no different from i.i.d. random variables in the sense that both satisfy a functional central limit theorem, while $I(1)$ processes are just like random walks in this sense. In practice, the dividing line between an persistent and non-persistent process is far less clear. Because cointegration is inherently about the low-frequency behavior of time series, persistence and non-persistence might more usefully be defined in terms of low frequency variability. This in turn requires a dividing line to define "low-frequencies", but natural definitions typically follow from the phenomenon under study. For example, in macroeconomics, long-run or low-frequency variability typically refers to frequencies lower than the business cycle, which are reasonably characterized by periodicities greater than 8 years. Thus, a macroeconomic time series might usefully be defined as $I(0)$ or "non-persistent" if it behaves like an i.i.d. process *for frequencies with periods longer than 8 years*, and otherwise it is "persistent". Müller and Watson (2007) use this idea to study univariate properties of economic time series, but the reasoning is equally (or more)

---

[1] See, for instance, Johansen (1988), Phillips and Hansen (1990), Saikkonen (1991), Park (1992) and Stock and Watson (1993).

compelling for cointegration.

As shown by Müller and Watson (2007), low-frequency variability can be summarized by a small number of weighted averages of the data, where the weights are low frequency trigonometric series. For example, only $q = 12$ weighted averages are needed to capture variability lower than the business cycle for time series that span 50 years (postwar data) regardless of the sampling frequency (months, quarters, weeks, etc.). Section 3 thus considers the behavior of these weighted averages as the sample size T grows large, but with q held fixed.[2] As in Bierens (1997), the weighted averages have a multivariate normal limiting distribution, and cointegration imposes restrictions on the covariance matrix of this distribution. Asymptotically, inference about the cointegrating vector thus becomes inference about the covariance matrix of a multivariate normal random vector. The low-frequency transformation approach is attractive beyond its statistical convenience because it explicitly acknowledges the relative scarcity of low-frequency information, it is robust to dynamic properties beyond the chosen frequency band, it does not require hard-to-interpret bandwidth choices, it is stable under aggregation, and it arguably gives the concept of "persistence" of an economic time series a straightforward interpretation.[3]

The second important issue in cointegration analysis involves the uncertain nature of the common stochastic trend. Elliott (1998) provides a dramatic demonstration of the fragility of standard cointegration methods by showing that they fail to control size when the common stochastic trends are not $I(1)$, but rather are "local-to-unity" in the sense of Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987) and Phillips (1987).[4] The development of valid tests for local-to-unity stochastic trends is complicated by the fact that the local-to-unity nuisance parameters cannot be consistently estimated. In a bivariate model, Cavanagh, Elliott, and Stock (1995) propose several procedures to adjust critical values from standard tests to control size over a range of values of the local-to-unity parameter, and their general

---

[2]This is discussed in further detail in Müller and Watson (2007); also see Bierens (1997), Phillips (1998) and Müller (2007b) who consider time series inference based on a finite number of weighted averages.

[3]An alternative to this low-frequency transformation approach is to model the persistence in the error correction term. A well-developed body of work has pursued this approach in the fractional integration framework, where the error correction term is allowed to have long memory. See, for instance, Jeganathan (1999), Kim and Phillips (2000), Robinson and Hualde (2003), Robinson and Marinucci (2003) and Velasco (2003); Canjels (1997) considers the same idea with a local-to-unity specification of the error correction term.

[4]Also see Elliott and Stock (1994) and Jeganathan (1997).

approach has been used by several other researchers; Campbell and Yogo (2006) provides a recent example. Stock and Watson (1996) and Jansson and Moreira (2006) go further and develop inference procedures with specific optimality properties in the local-to-unity model. In the fractional cointegration literature, the common stochastic trends are modelled as fractionally integrated, although the problem is different from the local-to-unity case as the fractional parameter can be consistently estimated under standard asymptotics. Yet, Müller and Watson (2007) demonstrate that, at least based on below business cycle variation, it is a hopeless endeavor to try to consistently discriminate between, say, local-to-unity and fractionally integrated stochastic data spanning 50 years.

As demonstrated by Wright (2000), it is nevertheless possible to conduct inference about a cointegrating vector without knowledge about the precise nature of the common stochastic trend. Wright's idea is to use the $I(0)$ property of the error correction term as the identifying property of the true cointegrating vector, so that a stationarity test of the model's putative error correction term is used to conduct inference about the value of the cointegrating vectors. Because the common stochastic trends drop out under the null hypothesis, Wright's procedure is robust in the sense that it controls size under any model for the common stochastic trend. But the procedure ignores the data beyond the putative error correction term, and is thus potentially quite inefficient.

Section 2 of this paper provides a formulation of the cointegrated model in which the common stochastic trends follow a flexible limiting Gaussian process that includes the $I(1)$, local-to-unity, and fractional/long-memory models as special cases. Throughout the paper, inference procedures are studied in the context of this general formulation of the cointegrated model. This may be viewed as a response to Granger's (1993) call to think of the persistence of macro time series as the result of a wide range of possible data generating processes beyond the $I(1)$ model, and to abandon attempts to identify the exact nature of the persistence process from the data.

The price to pay for this generality is that it introduces a potentially large number of nuisance parameters that complicate the derivation of efficient inference procedures. The nuisance parameters characterize the properties of the stochastic trends and the relationship between the stochastic trends and the model's $I(0)$ components. None of these nuisance parameters can be consistently estimated in the low-frequency framework outlined in Section 3, because they only affect the covariance matrix of the Gaussian limiting distribution of

3

the weighted averages. Invariance considerations makes a subset of these parameters irrelevant for the testing problem, but potentially many nuisance parameters remain. The main challenge of this paper is thus to study efficient tests in the presence of nuisance parameters under the null hypothesis, and Sections 4–6 address this issue.

Section 4 builds on Wright's (2000) suggestion and derives a low-frequency test for the value of the cointegration vectors based on an $I(0)$ test for the putative error correction term. Specifically, we derive a low-frequency version of a multivariate point-optimal scale and rotation invariant test against the alternative in which the common trends are $I(1)$. Similar to Wright's (2000) original suggestion, while simple, the application of this low-frequency test for inference about cointegrating vectors is potentially quite inefficient, as it ignores the data beyond the putative error correction term. But the null rejection probability of this test is unaffected by the properties of the common stochastic trend, so its power constitutes an easily achievable lower bound on the power of efficient tests.

Section 5 considers low-frequency tests based on all of the variables in the model. This section discusses an upper bound on the power of invariant tests of the value of the cointegrating vectors. The section discusses a general result concerning upper bounds for the power of tests when the null hypothesis involves nuisance parameters.[5] We then discuss numerical techniques to obtain low upper power bounds (approximate "least upper power bounds") for tests concerning the value of cointegrating vectors in our low-frequency framework. These bounds are computed for an alternative with the standard $I(1)$ stochastic trend, but under the constraint that the tests control size over a wide range of stochastic trend processes, ranging from the standard $I(1)$ model to a highly flexible model. These power bounds are useful for two main purposes. First, differences in the power bounds (interpreted as differences in least upper bounds) associated with restrictions on the trend process (for example, restricting the general stochastic trend process to be $I(1)$) quantify the restriction's information about the value of the cointegrating vector. Second, and most importantly, they provide a benchmark for the efficiency of any asymptotically valid test.

Section 6 combines the insights of Sections 4 and 5. In particular, the bounds derived in Section 5 allow us to assess the relative efficiency of the low-frequency version of Wright's (2000) test developed in Section 4. As it turns out, when attention is focused on a single

---

[5] The same insight about upper bounds on power was noted independently by Andrews, Moreira, and Stock (2007) and used for inference in IV models with potentially weak instruments.

cointegrating vector, and regardless of the number of common trends, the power of this test essentially coincides with the upper bound for an unrestricted version of the common trend process under the null hypothesis, and is close to the bound for several restricted, but still flexible common trend processes. Thus in this case, the low-frequency version of Wright's test—that is, ignoring the data beyond the putative error correction term—yields an essentially efficient test in the absence of strong *a priori* knowledge about the nature of the persistence.

One of the algorithms developed in Section 5 jointly determines a low power bound and a feasible test whose power is within a predetermined distance from the bound. The algorithm is generically applicable when the number of nuisance parameters is small and the density under the null and alternative is readily computable. Problems of this nature occur frequently in non-standard testing problems, so that this aspect of the present paper is likely to be of independent interest.

2 Model

Let p_t, $t = 1, ..., T$ denote the $n \times 1$ vector of variables under study. This section outlines a time domain representation of the cointegrated model for p_t in terms of canonical variables representing a set of common trends and $I(0)$ error correction terms. The common trends are allowed to follow a flexible process that includes $I(1)$, local-to-unity, and fraction models as special cases, but aside from this generalization, the cointegrated model for p_t is standard.

To begin, p_t is transformed into two components, where one component is $I(0)$ under the null hypothesis and the other component contains elements that are not cointegrated. Thus, let β denote an $n \times r$ matrix whose linearly independent columns are the cointegrating vectors, let β_0 denote the value of β under the null, and $y_t = \beta_0' p_t$. The elements in y_t are the model's error correction terms under the null hypothesis. Let $x_t = \delta' p_t$ where δ is $n \times k$ with $k = n - r$, and where the linearly independent columns of δ are linearly independent of the columns of β_0, so that the elements of x_t are not cointegrated under the null. Because the cointegrated model only determines the column space of the matrix of cointegrating vectors, the variables y_t and x_t are determined up to transformations $(y_t, x_t) \rightarrow (A_{yy} y_t, A_{xx} x_t + A_{xy} y_t)$, where A_{yy} and A_{xx} are non-singular. Most extant inference procedures are invariant (or asymptotically invariant) to these transformations, and, as discussed in detail below, our analysis will also

focus on invariant tests.

2.1 Canonical Variable Representation of y_t and x_t

We will represent y_t and x_t in terms of a common stochastic trend vector v_t and an $I(0)$ vector z_t

$$
\begin{aligned}
y_t &= \Gamma_{yz}z_t + \Gamma_{yv}v_t \\
x_t &= \Gamma_{xz}z_t + \Gamma_{xv}v_t,
\end{aligned} \tag{1}
$$

where z_t is $r \times 1$, v_t is $k \times 1$, and Γ_{yz} and Γ_{xv} have full rank. In this representation, the restriction that y_t is $I(0)$ corresponds to the restriction $\Gamma_{yv} = 0$. All of the test statistics discussed in this paper are invariant to adding constants to the observations, so that constant terms are suppressed in (1). Deterministic time trends will be briefly discussed in Section 5. As a technical matter, we think of $\{z_t, v_t\}_{t=1}^T$ (and thus also $\{x_t, y_t\}_{t=1}^T$) as being generated from a triangular array; we omit the additional dependence on T to ease notation. Also, we write $\lfloor x \rfloor$ for the integer part of $x \in \mathbb{R}$, $||A|| = \sqrt{\operatorname{tr} A'A}$ for any real matrix A, $x \vee y$ for the maximum of $x, y \in \mathbb{R}$, '$\otimes$' for the usual Kronecker product and '$\Rightarrow$' to indicate weak convergence.

Let $W(\cdot)$ denote a $n \times 1$ standard Wiener process. The vector z_t is a canonical $I(0)$ vector in the sense that its partial sums converge to a $r \times 1$ Wiener process

$$
T^{-1/2} \sum_{t=1}^{\lfloor sT \rfloor} z_t \Rightarrow S_z W(s) = W_z(s), \quad \text{where } S_z S_z' = I_r. \tag{2}
$$

The vector v_t is a common trend in the sense that scaled versions of its level converge to a stochastic integral with respect to $W(\cdot)$. For example, in the standard $I(1)$ model, $T^{-1/2}v_{\lfloor sT \rfloor} \Rightarrow \int_0^s H dW(t)$, where H is a $k \times n$ matrix and (H', S_z') has full rank. More general trend processes, such as the local-to-unity formulation, allow the matrix H to depend on s and t. The general representation for the common trends used in this paper is

$$
T^{-1/2}v_{\lfloor sT \rfloor} \Rightarrow \int_{-\infty}^s H(s,t)dW(t) \tag{3}
$$

where $H(s,t)$ is sufficiently well behaved to ensure that there exists a cadlag version of the process $\int_{-\infty}^s H(s,t)dW(t)$.[6]

---

[6]The common scale $T^{-1/2}$ for the $k \times 1$ vector v_t in (3) is assumed for convenience; with an appropriate

2.2 Invariance and Reparameterization

As discussed above, because cointegration only identifies the column space of β, attention is restricted to tests that are invariant to the group of transformations

$$(y_t, x_t) \rightarrow (A_{yy}y_t, A_{xx}x_t + A_{xy}y_t) \tag{4}$$

where A_{yy} and A_{xx} are non-singular, but (A_{yy}, A_{xx}, A_{xy}) are otherwise unrestricted real matrices.

The restriction to invariant tests allows a simplification of notation: because the test statistics are invariant to the transformations in (4), there is no loss of generality setting $\Gamma_{yz} = I_r$, $\Gamma_{xv} = I_k$, and $\Gamma_{xz} = 0$. With these values, the model is

$$\begin{aligned} y_t &= z_t + \Gamma_{yv}v_t \\ x_t &= v_t. \end{aligned} \tag{5}$$

2.3 Restricted Versions of the Trend Model

We will refer to the general trend specification in (3) as the "unrestricted" stochastic trend model throughout the remainder of the paper. The existing literature on efficient tests relies on restricted forms of the trend process (3) such as $I(1)$ or local-to-unity processes, and we compute the potential power gains associated with these and other *a priori* restrictions on $H(s,t)$ below. Here we introduce notation that is useful to describe five restricted versions of the stochastic trend.

The first model, which we will refer to as the G-model, restricts $H(s,t)$ to satisfy

$$H(s,t) = G(s,t)S_v, \tag{6}$$

where $G(s,t)$ is $k \times k$ and S_v is $k \times n$ with $S_vS_v' = I_k$ and (S_z', S_v') nonsingular. In this model, the common trend depends on $W(\cdot)$ only through the $k \times 1$ standard Wiener process $W_v(\cdot) = S_vW(\cdot)$, and this restricts the way that v_t and z_t interact. In this model

$$T^{-1/2}v_{\lfloor sT \rfloor} \Rightarrow \int_{-\infty}^{s} G(s,t)dW_v(t), \tag{7}$$

definition of local alternatives, the invariance (4) ensures that one would obtain the same results for any scaling of v_t.

7

and the covariance between the Wiener process characterizing the partial sums of z_t, W_z, and W_v is equal to the $r \times k$ matrix $R = S_z S_v'$. Standard $I(1)$ and local-to-unity formulations of cointegration satisfy this restriction and impose additional parametric restrictions on $G(s,t)$.

The second model further restricts (7) so that $G(s,t)$ is diagonal:

$$G(s,t) = \mathrm{diag}(g_1(s,t), \cdots, g_k(s,t)). \tag{8}$$

An interpretation of this model is that the k common trends evolve independently of one another (recall that W_v has identity covariance matrix), where each trend is allowed to follow a different process characterized by the functions $g_i(s,t)$.

The third model further restricts the diagonal-G model so that the k stochastic trends converge weakly to a stationary continuous time process. We thus impose

$$g_i(s,t) = g_i^S(s-t), \; i = 1, \cdots, k. \tag{9}$$

The stationary local-to-unity model (with an initial condition drawn from the unconditional distribution), for instance, satisfies this restriction.

Finally, we consider two parametric restrictions of G:

$$G(s,t) = \mathbf{1}[t > 0]I_k \tag{10}$$

which is the $I(1)$ model, and

$$G(s,t) = \mathbf{1}[t > 0]e^{C(s-t)} \tag{11}$$

which is the multivariate local-to-unity model, where C is the $k \times k$ diffusion matrix of the limiting Ornstein-Uhlenbeck process (with zero initial condition).

Because of the invariance (4), the trend models are unaffected by premultiplication of $H(s,t)$ (or $G(s,t)$) by an arbitrary non-singular $k \times k$ matrix. (This is why the $I(1)$ specification in (10) is the same as the $I(1)$ specification given below (2).)

2.4   Testing Problem and Local Alternatives

The goal of the paper is to derive asymptotically efficient tests for the value of the cointegrating vectors with controlled rejection probability under the null hypothesis for a range of stochastic trend specifications. The different orders of magnitude of z_t and v_t in (2) and (3)

suggest a local embedding of this null hypothesis against alternatives where $\Gamma_{yv} = T^{-1}B$ for B a constant $r \times k$ matrix, so that in model (5),

$$T^{-1/2} \sum_{t=1}^{\lfloor sT \rfloor} y_t \Rightarrow S_z W(s) + B \int_0^s \int_{-\infty}^u H(u,t)dW(t)du.$$

In this parametrization, the null hypothesis becomes

$$H_0 : B = 0, \quad H(s,t) \in \mathcal{H}_0 \tag{12}$$

where $H(s,t)$ is restricted to a set of functions $\mathcal{H}_0$, that, in the unrestricted trend model includes functions sufficiently well behaved to ensure that there exists a cadlag version of the process $\int_{-\infty}^s H(s,t)dW(t)$, or more restricted versions of $H(s,t)$ as in (6), (8), (9), (10), or (11).

Since our goal is to consider efficient tests of the null hypothesis (12), we also need to specify the alternative hypothesis. Our results below are general enough to allow for the derivation of efficient tests against any particular alternative with specified $B = B_1$ and stochastic trend process $H(s,t) = H_1(s,t)$,

$$H_a : B = B_1, \quad H(s,t) = H_1(s,t) \tag{13}$$

or, more generally, for tests that are efficient in the sense of maximizing weighted average power against a set of values for B_1 and stochastic trend models $H_1(s,t)$.

Our numerical results, however, focus on alternatives in which the stochastic trend v_t is $I(1)$, so that $H_1(s,t)$ satisfies (6) and (10). This is partly out of practical considerations: while there is a wide range of potentially interesting trend specification, the computations for any particular specification are involved, and these computational complications limit the number of alternatives we can usefully consider. At the same time, one might consider the classical $I(1)$ model as an important benchmark against which it is useful to maximize power—not necessarily because this is the only plausible model under the alternative, but because a test that performs well against this alternative presumably has reasonable power properties for a range of empirically relevant models. We stress that despite this focus on the $I(1)$ stochastic trend model for the alternative hypothesis (13), we restrict attention to tests that control size for a range of models under the null hypothesis (12). The idea is to control the frequency of rejections under the null hypothesis for any stochastic trend model

9

in $\mathcal{H}_0$, so that the rejection of a set of cointegrating vectors cannot simply be explained by the stochastic trends not being exactly $I(1)$. In this sense, our approach is one of "robust" cointegration testing, with the degree of robustness governed by the size of the set $\mathcal{H}_0$.

2.5   Summary

To summarize, this section has introduced the time domain representation of the cointegrated model with a focus on the problem of inference about the space of cointegrating vectors. In all respects except one, the representation is the standard one: the data are expressed as a linear function of a canonical vector or common trends and a vector of $I(0)$ components. Under the null, certain linear combinations of the data do not involve the common trends. Because the null only restricts the column space of the matrix of cointegrating vectors, attention is restricted to invariant tests. The goal is to construct tests with best power for an alternative value for the matrix of cointegrating vectors under a particular model for the trend (or best weighted average power for a collection of B_1 and $H_1(s, t)$). The formulation differs from the standard one only in that it allows the model for trend under the null to be less restrictive than the standard formulation. Said differently, because of potential uncertainty about the specific form of the trend process, the formulation restricts attention to tests that control size for a range of different trend processes. As will become evident, this generalization complicates the problem of constructing efficient tests by introducing a potentially large number of nuisance parameters (associated with trend process) under the null hypothesis.

3   Low-Frequency Representation of the Model

Cointegration is a restriction on the low-frequency behavior of time series, and as discussed in the introduction, we therefore focus on the low-frequency behavior of (y_t, x_t). This low-frequency variability is summarized by a small number, q, of weighted averages of the data. In this section we discuss these weighted averages and derive their limiting behavior under the null and alternative hypotheses.

3.1 Low-Frequency Weighted Averages

We use weights associated with the cosine transform, where the j'th weight is given by $\Psi_j(s) = \sqrt{2}\cos(j\pi s)$. For any sequence $\{a_t\}_{t=1}^T$, the j'th weighted average will be denoted by

$$A_T(j) = \int_0^1 \Psi_j(s)a_{\lfloor sT\rfloor+1}ds = \iota_{jT}T^{-1}\sum_{t=1}^T \Psi_j(\tfrac{t-1/2}{T})a_t \tag{14}$$

where $\iota_{jT} = (2T/j\pi)\sin(j\pi/2T) \to 1$ for all fixed j. As demonstrated by Müller and Watson (2007), the weighted averages $A_T(j)$, $j = 0, \cdots, q$, essentially capture the variability in the sequence corresponding to frequencies below $q\pi/T$.

We use the following notation: with a_t a $h \times 1$ vector time series, let $\Psi(s) = (\Psi_1(s), \Psi_2(s), \cdots, \Psi_q(s))'$ denote the $q \times 1$ vector of weighting functions, and $A_T = \int_0^1 \Psi(s)a'_{\lfloor sT\rfloor+1}ds$ the $q \times h$ matrix of weighted averages of the elements of a_t, where $\Psi_0(s)$ is excluded to make the results invariant to adding constants to the data. Using this notation, the $q \times r$ matrix Y_T and the $q \times k$ matrix X_T summarize the variability in the data corresponding to frequencies lower than $q\pi/T$. With $q = 12$, (Y_T, X_T) capture variability lower than the business cycle (periodicities greater than 8 years) for time series that span 50 years (postwar data) regardless of the sampling frequency (months, quarters, weeks, etc.). This motivates us to consider the behavior of these matrices as $T \to \infty$, but with q held fixed.

The large-sample behavior of X_T and Y_T follows from the behavior of Z_T and V_T. Using the assumed limits (2) and (3), the continuous mapping theorem, and integration by parts for the terms involves Z_T, one obtains

$$\begin{bmatrix} T^{1/2}Z_T \\ T^{-1/2}V_T \end{bmatrix} \Rightarrow \begin{bmatrix} Z \\ V \end{bmatrix} \tag{15}$$

where

$$\begin{bmatrix} \mathrm{vec}\,Z \\ \mathrm{vec}\,V \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} I_{rq} & \Sigma_{ZV} \\ \Sigma_{VZ} & \Sigma_{VV} \end{bmatrix}\right) \tag{16}$$

with

$$\Sigma_{VZ} = \int_0^1 \left(\int_{t\vee 0}^1 [H(s,t) \otimes \Psi(s)]ds\right)[S_z \otimes \Psi(t)]'dt \tag{17}$$

$$\Sigma_{VV} = \int_{-\infty}^1 \left(\int_{t\vee 0}^1 [H(s,t) \otimes \Psi(s)]ds\right)\left(\int_{t\vee 0}^1 [H(s,t) \otimes \Psi(s)]ds\right)'dt.$$

11

The relative scarcity of low-frequency information is thus formally captured by considering the weak limits (2) and (3) as pertinent only for the subspace spanned by the weight function $\Psi(\cdot)$, yielding (15) as a complete characterization of the relevant properties of the error correction term z_t and the common stochastic trend v_t.

Using $\Gamma_{yv} = T^{-1}B$, equation (5) implies that $Y_T = Z_T + T^{-1}V_T B'$ and $X_T = V_T$. Thus,

$$\begin{bmatrix} T^{1/2}Y_T \\ T^{-1/2}X_T \end{bmatrix} \Rightarrow \begin{bmatrix} Y \\ X \end{bmatrix} = \begin{bmatrix} Z + VB' \\ V \end{bmatrix} \tag{18}$$

where

$$\begin{bmatrix} \text{vec}\,Y \\ \text{vec}\,X \end{bmatrix} \sim \mathcal{N}\left(0, \Sigma_{(Y,X)}\right) \tag{19}$$

with

$$\Sigma_{(Y,X)} = \begin{bmatrix} I_r \otimes I_q & B \otimes I_q \\ 0 & I_k \otimes I_q \end{bmatrix} \begin{bmatrix} I_r \otimes I_q & \Sigma_{ZV} \\ \Sigma_{VZ} & \Sigma_{VV} \end{bmatrix} \begin{bmatrix} I_r \otimes I_q & 0 \\ B' \otimes I_q & I_k \otimes I_q \end{bmatrix}. \tag{20}$$

3.2 "Best" Low-Frequency Hypothesis Tests

We consider invariant tests of H_0 against H_a given in (12) and (13) based on the data $\{y_t, x_t\}_{t=1}^{T}$. Because we are concerned with the model's implications for the low-frequency variability of the data, we restrict attention to tests that control asymptotic size for all models that satisfy (18)-(20). Our goal is to find an invariant test that maximizes power subject to this restriction, and for brevity we will refer to such a test as a "best" test. Müller (2007a) considers the general problem of constructing asymptotically most powerful tests subject to asymptotic size control over a class of models such as ours. In our context, his results imply that asymptotically best tests correspond to the most powerful invariant tests associated with the limiting distribution (19).

Thus, the relevant testing problem has a simple form: $\text{vec}(Y, X)$ has a normal distribution with mean zero and covariance matrix that depends on B. Under the null $B = 0$, while under the alternative $B \neq 0$. Tests are restricted to be invariant to the group of transformations

$$(Y, X) \rightarrow (Y A'_{yy}, X A'_{xx} + Y A'_{xy}) \tag{21}$$

where A_{yy} and A_{xx} are nonsingular, and A_{yy}, A_{xx}, and A_{xy} are otherwise unrestricted. Thus, the hypothesis testing problem becomes the problem of using an invariant procedure to test a restriction on the covariance matrix of a multivariate normal vector.

12

4   Efficient Y-Only Tests

The primary obstacle for constructing efficient tests of the null hypothesis that $B = 0$ is the large number of nuisance parameters associated with the stochastic trend (the parameters that determine $H(s,t)$). These parameters govern the values of Σ_{ZV} and Σ_{VV}, which in turn determine Σ_{YX} and Σ_{XX}. Any valid test must control size over all values of these nuisance parameters. Wright (2000) notes that this obstacle can be avoided by ignoring the x_t data and basing inference only on y_t, since under the null hypothesis, $y_t = z_t$. This section takes up Wright's suggestion and discusses efficient low-frequency "Y-only" tests.[7]

   The power of the resulting test serves as an easily achievable lower bound on the power of efficient tests that exploit both Y and X. The next section discusses upper bounds on the power of tests that rely on both Y and X, so that the relative efficiency of the Y-only test can be evaluated.

4.1   Efficient Tests against General Alternatives

The distribution of $\operatorname{vec} Y \sim \mathcal{N}(0, \Sigma_{YY})$ follows from the derivations in Section 3: Under the null hypothesis, $\Sigma_{YY} = I_{rq}$, and under the alternative, Σ_{YY} depends on the local alternative B, the properties of the stochastic trend and its relationship with the error correction term Z. For a particular choice of alternative, the testing problem thus becomes $H_0 : \Sigma_{YY} = I$ against $H_1 : \Sigma_{YY} = \Sigma_{YY1}$, and the invariance requirement (21) reduces to

$$Y \to Y A'_{yy} \quad \text{for arbitrary nonsingular } r \times r \text{ matrices } A_{yy}. \tag{22}$$

Any invariant test can be written as a function of a maximal invariant (Theorem 6.2.1 in Lehmann and Romano (2005)), so that by the Neyman-Pearson lemma, the most powerful invariant test rejects for large values of the likelihood ratio statistic of a maximal invariant.

---

[7] Wright (2000) implements this idea using a "stationarity" test of the $I(0)$ null proposed by Saikkonen and Luukonen (1993), using a robust covariance matrix as in Kwiatkowski, Phillips, Schmidt, and Shin (1992) for the test proposed in Nyblom (1989). This test relies on a consistent estimator of the spectral density matrix of z_t at frequency zero. But consistent estimation requires a lot of pertinent low frequency information, and lack thereof leads to well-known size control problems (see for example, Kwiatkowski, Phillips, Schmidt, and Shin (1992), Caner and Kilian (2001), and Müller (2005)). These problems are avoided by using the low-frequency components of y_t only; see Müller and Watson (2007) for further discussion.

The only remaining challenge is the computation of the density of a maximal invariant, and this is addressed in the following theorem.

Theorem 1 *(a) If* $\operatorname{vec} Y \sim \mathcal{N}(0, \Sigma_{YY})$ *with positive definite* Σ_{YY} *and* $q > r$, *the density of a maximal invariant to (22) has the form*

$$c_1 (\det \Sigma_{YY})^{-1/2} (\det \Omega_Y)^{-1/2} E_{\omega_Y} [|\det(\omega_Y)|^{q-r}]$$

where c_1 *does not depend on* Σ_{YY}, ω_Y *is an* $r \times r$ *random matrix with* $\operatorname{vec} \omega_Y \sim \mathcal{N}(0, \Omega_Y^{-1})$, $\Omega_Y = (I_r \otimes Y)' \Sigma_{YY}^{-1} (I_r \otimes Y)$, *and* E_{ω_Y} *denotes integration with respect to the distribution of* ω_Y *(conditional on* Y*).*

(b) If in addition, $\Sigma_{YY} = \tilde{V}_{YY} \otimes \tilde{\Sigma}_{YY}$, *then the density simplifies to*

$$c_2 (\det \tilde{\Sigma}_{YY})^{-r/2} \det(Y' \tilde{\Sigma}_{YY}^{-1} Y)^{-q/2}$$

where c_2 *does not depend on* Σ_{YY}.

Part (a) of the theorem provides a formula for the density of a maximal invariant in terms of absolute moments of the determinant of a multivariate normal matrix, whose covariance matrix depends on Y. We know of no useful closed-form solution for this expectation, however the density can be evaluated by appropriate Monte Carlo techniques. Part (b) provides an explicit and simple formula when the covariance matrix is of a specific Kronecker form. This form arises under the null hypothesis with $\Sigma_{YY} = I_{rk}$, and under alternatives where each of the r putative error correction terms in y_t have the same low-frequency covariance matrix. For a simple alternative hypothesis with $\Sigma_{YY1} = \tilde{V}_{YY1} \otimes \tilde{\Sigma}_{YY1}$, the best test then rejects for large values of $\det(Y'Y) / \det(Y' \tilde{\Sigma}_{YY1}^{-1} Y)$. The form of weighted average power maximizing tests over a set of alternative covariance matrices Σ_{YY1} are also easily deduced from Theorem 1 parts (a) and (b).

4.2   Efficient Tests against $I(1)$ Alternative

As discussed above, the numerical results in this paper focus on the benchmark alternative where the stochastic trend follows an $I(1)$. Under this alternative, y_t follows a multivariate "local level model" (cf. Harvey (1989)), which is the alternative underlying well-known "stationarity" tests such as Nyblom and Mäkeläinen (1983), Nyblom (1989), Kwiatkowski,

14

Phillips, Schmidt, and Shin (1992), Nyblom and Harvey (2000), Jansson (2004), and others. Thus, suppose that the stochastic trend satisfies (6) and (10), so that

$$T^{-1/2} \sum_{t=1}^{\lfloor sT \rfloor} y_t \Rightarrow W_z(s) + B \int_0^s W_v(t) dt. \tag{23}$$

The optimal test depends on the value of B under the alternative, and it is convenient to assume

$$B = bS, \tag{24}$$

where b is a scalar and S is the $r \times k$ selection matrix equal to $S = [I_r, 0_{k-r}]$ when $r \leq k$ and $S = [I_k, 0_{k-r}]'$ when $r > k$. The invariance requirement (4) implies that (24) is without loss of generality whenever there exist orthonormal $r \times r$ and $k \times k$ matrices P_y and P_x such that $P_y B P_x = ||B||S$, which is always the case when $\min(r, k) = 1$. In the formulation (24), when $r \leq k$ (so that the number of linearly independent cointegrating vectors does not exceed the number of common trends), each element of y_t is the sum of an $I(0)$ component and an $I(1)$ component, where the common relative magnitude of the two components is determined by b. When $r > k$, there are fewer trends than cointegrating vectors, so that y_t can be rotated such that the trends load on only a subset of the variables in y_t. This is the "reduced rank" formulation used, for example, in the multivariate stationarity test proposed in Eliasz, Stock, and Watson (2004).

In this model, the covariance matrix of Y depends on b and $R = S_z S_v' = E[W_z(1)W_v(1)']$, the correlation between the Wiener processes describing z_t and v_t. A straightforward calculation shows that Σ_{YY} can be written as

$$\Sigma_{YY} = (I_r \otimes I_q) + b^2(SS' \otimes D) + b(SR' \otimes F) + b(RS' \otimes F') \tag{25}$$

where F and D are $q \times q$ matrices, where D is a diagonal matrix with i'th diagonal element equal to $(\pi i)^{-2}$ and $F = [f_{ij}]$, with $f_{ij} = 0$ if i and j are both even or odd, and $f_{ij} = 4/[\pi^2(i^2 - j^2)]$ otherwise. (The simple diagonal form of D is due to the particular choice of the weighting functions Ψ in (14); see Section 2.3 in Müller and Watson (2007)).

Examination of (25) suggests three simplifications of the testing problem. First, because $F = -F'$, the final two terms cancel when SR' is symmetric. (When $r \leq k$, SR' is symmetric if $R_{ij} = R_{ji}$ for $i, j \leq r$, and when $r > k$, symmetry requires $R_{ij} = R_{ji}$ for $i, j \leq k$ and $R_{ij} = 0$ for $i > k$ and all j.) Thus, when SR' is symmetric, Σ_{YY} does not depend on R,

15

which implies that the efficient test constructed using $R = 0$ is uniformly most powerful for all values of R with SR' symmetric. Second, when $r \leq k$, $SS' = I_r$, so in this case when SR' is symmetric, $\Sigma_{YY} = I_r \otimes (I_q + b^2 D)$, and from part (b) of Theorem 1, the optimal test rejects for large values of $\det(Y'Y)/\det(Y'(I_q + b^2 D)^{-1} Y)$. (We will refer to this statistic as "JW" to highlight its origin in the ideas in Wright (2000).) Finally, when SR' is symmetric, but $r > k$, a calculation based on Theorem 1 (a) produces an expression for the best test. These simplifications are summarized in the following corollary.

Corollary 1 *For the alternative (23) and (24), the Neyman-Pearson test constructed with $R = 0$ is uniformly most powerful over all values of R with SR' symmetric, and rejects for large values of*

$$JW(b) = \det(Y'Y)/\det(Y'(I_q + b^2 D)^{-1} Y)$$

when $r \leq k$, and for large values of

$$\xi(b) = \det(Y'Y)^{(q+k-r)/2} \det(Y'(I_q + b^2 D)^{-1} Y)^{-k/2} E_{\omega_Y}[|\det(\omega_Y)|^{q-r}]$$

when $r > k$, where $\operatorname{vec} \omega_Y \sim \mathcal{N}(0, \Omega_Y^{-1})$ and $\Omega_Y = \operatorname{diag}(I_k \otimes Y'(I_q + b^2 D)^{-1} Y, I_{r-k} \otimes Y'Y)$.

Table 1 presents 10%, 5%, and 1% critical values for the point-optimal $JW(10/\sqrt{r})$ test for various values of r and q, where the alternative is chosen so that 5% test has approximately 50% power for $b = 10/\sqrt{r}$.

In many empirical applications, $r = 1$, so that the best test based on Y is the JW test using critical values given in the first column of the table. When $r > 1$, JW is the best test when $r \leq k$, but when $r > k$, the optimal test statistic is ξ given in part (b) of the corollary. This test statistic is more difficult to calculate than JW because ξ depends on the term $E_{\omega_Y}[|\det(\omega_Y)|^{q-r}]$, which requires evaluating absolute moments of order $q - r$ from an r^2-dimensional multivariate normal distribution. In a typical application $q = 12$ and even with $r = 2$ or 3, accurate Monte Carlo evaluation of $E[|\det(\omega_Y)|^{q-r}]$ requires a large number of draws, even using importance sampling methods. Because of the large marginal computational cost of ξ over JW, it is natural to consider the marginal gain in power associated with using ξ instead of JW. Figure 1 compares the power of $\xi(b)$ and $JW(b)$ for $(r, k) = (2, 1), (3, 1)$, and $(3, 2)$ in models with SR' symmetric for $q = 12$. The figures show the power bound associated with tests based on $\xi(b)$, and the power of the corresponding

16

$JW(b)$ for various values of b. When $r = 2$ and $k = 1$, the power of the $JW(b)$ statistic is within 3% of the power of $\xi(b)$ when the power of $\xi(b)$ is less than 50%, but the difference increases to nearly 10% when the power of $\xi(b)$ exceeds 80%. The differences are more substantial when $r = 3$ and $k = 1$, where the power difference is approximately 7% when power is 50%; the power differences are negligible when $r = 3$ and $k = 2$. Eliasz, Stock, and Watson (2004) report similar power differences in a related testing problem.

5 Power Bounds for Tests Using (Y, X)

The tests proposed in the last section avoided the complications associated with the nuisance parameters describing the commons stochastic trend model by ignoring the information in X and basing inference only on Y. This section considers invariant tests that utilize information in both Y and X.

The analysis proceeds in four steps. The first step characterizes the distribution of the maximal invariant of (Y, X), which forms the basis for efficient invariant inference. The second step determines convenient parameterizations of the covariance matrix $\Sigma_{(Y,X)}$ in (20) under the null and alternative hypotheses. A third step presents a general result about power bounds for testing problems involving nuisance parameters under the null hypothesis, which is closely related to the concept of a least favorable distribution for the nuisance parameters. Finally, the fourth step discusses numerical techniques to implement the power bounds for the problem under study. Each of these steps is presented in a subsection.

The penultimate subsection shows that power bounds for $r = 1$ and $k = 2$ also hold for models with $r = 1$ and $k > 2$, and a final subsection offers a brief discussion of how linear trends would affect the analysis.

5.1 Density of a Maximal Invariant

Theorem 1 presented the marginal density of a maximal invariant for $Y \rightarrow Y A'_{yy}$. The following theorem provides an expression for the density of a maximal invariant for $(Y, X) \rightarrow (Y A'_{yy}, X A'_{xx} + Y A'_{xy})$.

Theorem 2 *If* $\text{vec}(Y,X) \sim \mathcal{N}(0,\Sigma_{(Y,X)})$ *with positive definite* $\Sigma_{(Y,X)}$ *and* $q > r + k$, *the density of a maximal invariant of (21) has the form*

$$c(\det \Sigma_{(Y,X)})^{-1/2}(\det V'_{0Y}\Sigma^{-1}_{(Y,X)}V_{0Y})^{-1/2}(\det \Omega)^{-1/2}E_\omega[|\det(\omega_Y)|^{q-r}|\det(\omega_X)|^{q-r-k}]$$

where c *does not depend on* $\Sigma_{(Y,X)}$, ω_Y *and* ω_X *are random* $r \times r$ *and* $k \times k$ *matrices, respectively, with* $(\text{vec}\,\omega'_y, \text{vec}\,\omega'_x)' \sim \mathcal{N}(0,\Omega^{-1})$,

$$\Omega = D'_{YX}\Sigma^{-1}_{(Y,X)}D_{YX} - D'_{YX}\Sigma^{-1}_{(Y,X)}V_{0Y}(V'_{0Y}\Sigma^{-1}_{(Y,X)}V_{0Y})^{-1}V'_{0Y}\Sigma^{-1}_{(Y,X)}D_{YX},$$

$D_{YX} = \text{diag}(I_r \otimes Y, I_k \otimes X)$, $V_{0Y} = (0'_{rq \times rk}, I_k \otimes Y')'$, *and* E_ω *denotes integration with respect to* ω_Y *and* ω_X, *conditional on* (Y,X).

Similar to Theorem 1, Theorem 2 shows that density of a maximal invariant can be expressed in terms of absolute moments of determinants of jointly normally distributed random matrices, whose covariance matrix depends on (Y,X). We do not know of a useful and general closed-form solution for this expectation; for $r = k = 1$, however, Nabeya's (1951) results for the absolute moments of a bivariate normal yields an expression in terms of elementary functions, which we omit for brevity. When $r + k > 2$ and $q \approx 12$, one can compute the moments via Monte Carlo integration. However, computing accurate approximations is difficult when r and k are large, and the numerical analysis reported below is therefore limited to small values of r and k.

5.2   Parameterization of $\Sigma_{(Y,X)}$

Since the density of the maximal invariant of Theorem 2 depends on $\Sigma_{(Y,X)}$, the derivation of efficient invariant tests requires specification of $\Sigma_{(Y,X)}$ under the alternative and null hypothesis. We discuss each of these in turn.

5.2.1   Specification of $\Sigma_{(Y,X)}$ under the Alternative Hypothesis

As discussed above, we focus on the alternative where the stochastic trends follow an $I(1)$ process, so that $H(s,t)$ satisfies (6) and (10). There remains the issue of the value of B (the coefficients that show how the trends affect Y) and R (the correlation of the Wiener processes describing the $I(0)$ variables, z_t, and the common trends, v_t). For these parameters,

we consider point-valued alternatives with $B = B_1$ and $R = R_1$; the power bounds derived below then serve as bounds on the asymptotic power envelope over these values of B and R. Invariance reduces the effective dimension of B and R somewhat, and this will be discussed in the context of the numerical results presented below.

5.2.2 Parameterization of $\Sigma_{(Y,X)}$ under the Null Hypothesis

From (20), under the null hypothesis with $B = 0$, the covariance matrix $\Sigma_{(Y,X)}$ satisfies

$$\Sigma_{(Y,X)} = \begin{bmatrix} I_{rq} & \Sigma_{ZV} \\ \Sigma_{VZ} & \Sigma_{VV} \end{bmatrix}.$$

The null model's specification of the stochastic trend determines the $rq \times kq$ matrix Σ_{ZV} and the $kq \times kq$ matrix Σ_{VV} by the formulae given in (17). Since these matrices contain a finite number of elements, it is clear that even for nonparametric specifications of $H(s,t)$, the effective parameter space for low-frequency tests based on (Y, X) is finite dimensional. We collect these nuisance parameters in a vector $\theta \in \Theta$.

We now discuss convenient parameterizations of $\Sigma_{(Y,X)}$ for the unrestricted stochastic trend model, and the five restrictions discussed in Section 2: the first restricts $H(s,t) = G(s,t)S_v$ as in (6), the second further restricts $G(s,t)$ to be diagonal, and the third imposes a stationarity restriction on the diagonal model. The final two models impose the local-to-unity and $I(1)$ parametric restrictions on the trend process.

The following lemma provides the basis for parameterizing $\Sigma_{(Y,X)}$ when $H(s,t)$ is unrestricted.

Lemma 1 (a) For any $(r+k)q \times (r+k)q$ positive definite matrix Σ^* with upper left $rq \times rq$ block equal to I_{rq}, there exists an unrestricted trend model with $H(s,t) = 0$ for $t < 0$ such that $\Sigma^* = E[\text{vec}(Z,V)(\text{vec}(Z,V))']$.

(b) If $r \leq k$, this $H(s,t)$ can be chosen of the form $H(s,t) = G(s,t)S_v$, where (S_z', S_v') has full rank.

Thus, when $H(s,t)$ is unrestricted or $r \leq k$ and $H(s,t) = G(s,t)S_v$, the only restriction that the null hypothesis imposes on $\Sigma_{(Y,X)}$ is that $\Sigma_{YY} = I_{rq}$. In other words, since Σ_{ZV} and Σ_{VV} have $rkq^2 + kq(kq+1)/2$ distinct elements, an appropriately chosen θ of that dimension

determines $\Sigma_{(Y,X)}$ under the null hypothesis in the unrestricted model, and in the model where $H(s,t) = G(s,t)S_v$ for $r \leq k$.

When $H(s,t) = G(s,t)S_v$ and $r > k$, and when $G(s,t) = \text{diag}(g_1(s,t), \cdots, g_k(s,t))$, there are additional restrictions on $\Sigma_{(Y,X)}$, which are most easily parametrized be taking invariance considerations into account. We discuss these restrictions in the appendix. In the stationary diagonal model where $G(s,t) = \text{diag}(g_1^S(s-t), \cdots, g_k^S(s-t))$, we set g_j^S to be step functions

$$g_j^S(x) = \sum_{i=1}^{n_g} \xi_{j,i} \mathbf{1} \left[\frac{i-1}{n_g+1} \leq \frac{x}{1+x} < \frac{i}{n_g+1} \right] \tag{26}$$

for $n_g = 40$ and some scalar parameters $\xi_{j,i}$, $j = 1, \cdots, k$, $i = 1, \cdots, n_g$. The steps occur at the points $i/(n_g+1-i)$, so that more flexibility is allowed for small values of x (half of the points are associated with values of x less than 1, for example). The values of Σ_{ZV} and Σ_{VV} then follow from (17). In this specification θ contains the kn_g coefficients $\xi_{j,i}$ and the rk coefficients in the correlation matrix R. While the specification (26) only captures a subset of all possible covariance matrices $\Sigma_{(Y,X)}$ in the (nonparametric) stationary model, any test that controls size for all functions $H(s,t)$ of the form $H(s,t) = \text{diag}(g_1^S(s-t), \cdots, g_k^S(s-t))S_v$ a fortiori has to control size for the specification (26). The upper bounds on power of tests that control size for all values of θ under (26) are therefore also upper bounds for tests that control size for the generic stationary stochastic trend model.

In the local-to-unity model, a straightforward (but tedious) calculation determines the value of $\Sigma_{(Y,X)}$ as function of the $r \times r$ matrix C and the $r \times k$ correlation matrix R, so that θ is of dimension $r^2 + rk$. Finally, the $I(1)$ model is a special case of the local-to-unity model with $C = 0$.

5.3 A General Result about Power Bounds

The general version of the hypothesis testing problem we are facing is a familiar one: Let U denote a single observation of dimension $m \times 1$. (In our problem, the maximal invariant discussed in Theorem 2 plays the role of U.) Under the null hypothesis U has probability density $f_\theta(u)$ with respect to some measure μ, where $\theta \in \Theta$ is a vector of nuisance parameters. (In our problem, the vector θ describes the stochastic trend process under the null hypothesis and determines $\Sigma_{(Y,X)}$ via (17) and (20) as discussed in the last subsection). Under the

alternative, U has known density $h(u)$. (Choices for $h(u)$ were discussed in Subsection 5.2.1.) Thus, the null and alternative hypothesis are

$$H_0 : \text{The density of } U \text{ is } f_\theta(u), \ \theta \in \Theta$$
$$H_1 : \text{The density of } U \text{ is } h(u), \tag{27}$$

and possibly randomized tests are (measurable) functions $\varphi : \mathbb{R}^m \mapsto [0,1]$, where $\varphi(u)$ is the probability of rejecting the null hypothesis when observing $U = u$, so that size and power are given by $\sup_{\theta \in \Theta} \int \varphi f_\theta d\mu$ and $\int \varphi h d\mu$, respectively. The aim is to construct an efficient test φ^*. Unfortunately, there does not exist a general method to construct such an efficient test φ^* in the presence of the nuisance parameter θ.

Thus, to make further progress, this subsection shows how to derive a set of upper bounds on the power of all level α tests of (27). In general, an upper power bound may be useful for at least two purposes: First, one might know that some *ad hoc* test is of level α (i.e. its size is at most α). If the power of this *ad hoc* test happens to be close to the upper bound, then one can conclude that the *ad hoc* test is close to efficient, since by definition of the upper bound, no test exists with higher power. (This reasoning will be used to argue for the approximate efficiency of the JW statistic derived in Section 4 in the unrestricted stochastic trend model). Second, when the dimension of θ is small, one can numerically search for a level α test with high power. An upper bound on the power of all level α tests is useful in this approach, because if the numerical search has identified a test with almost as much power as the upper bound, one has identified a test that is efficient for all practical purposes. Subsection 5.4.1 below discusses such an algorithm.

A standard device for problems such as (27) is to consider a Neyman-Pearson test for a related problem in which the null hypothesis is replaced with a mixture

$$H_\Lambda : \text{The density of } U \text{ is } \int f_\theta d\Lambda(\theta)$$

where Λ is a probability distribution for θ. The following lemma shows that the power of the Neyman-Pearson test of H_Λ versus H_1 provides an upper power bound for tests of H_0 versus H_1.

Lemma 2 *Let φ_Λ be the best level α test of H_Λ against H_1. Then for any level α test φ of H_0 against H_1, $\int \varphi_\Lambda h d\mu \geq \int \varphi h d\mu$.*

21

Proof. *Since φ is a level α test of H_0, $\int \varphi f_\theta d\mu \leq \alpha$ for all $\theta \in \Theta$. Therefore, $\int \int \varphi f_\theta d\mu d\Lambda(\theta) = \int \int \varphi f_\theta d\Lambda(\theta) d\mu \leq \alpha$ (where the change in the order of integration is allowed by Fubini's Theorem), so that φ is also a level α test of H_Λ against H_1. The result follows by the definition of a best test.* ∎

This result is closely related to Theorem 3.8.1 of Lehmann and Romano (2005) which provides conditions under which a least upper bound on the power for tests H_0 versus H_1 is associated with a "least favorable distribution" for θ, and that using this distribution for Λ produces the least upper power bound. The least favorable distribution Λ^* has the characteristic that the resulting φ_{Λ^*} is a level α test for testing H_0 versus H_1. Said differently, if φ_{Λ^*} is the best level α test of H_{Λ^*} against H_1 *and* is also a level α test for testing H_0 versus H_1, then $\varphi^* = \varphi_{\Lambda^*}$, that is φ_{Λ^*} is the most powerful level α test of H_0 versus H_1. Unfortunately, while the test associated with the least favorable distribution solves the testing problem (27), there is no general and constructive method for finding the least favorable distribution Λ^* (and it does not always exist).

With this in mind, Lemma 2 is stated so that Λ is not necessarily the least favorable distribution. That is, the bound in Lemma 2 holds for any probability distribution Λ. The goal of the numerical analysis carried out below is to choose Λ to approximate the least upper bound. Importantly, even if one cannot identify the least favorable distribution, Lemma 2 shows that the power of φ_Λ provides a valid bound for the power of any test of H_0 versus H_1, for any Λ.

5.4 Numerical Implementation

This paper relies on Lemma 2 in two distinct ways. On the one hand, we develop an algorithm that simultaneously determines a low upper bound on power, and a level α test whose power is close to that bound. This algorithm is entirely generic in the sense that it does not exploit any specificities of the low-frequency robust cointegration testing problem; in practice, it only requires that the densities f_θ and h can be quickly evaluated numerically. The computational complexity is such, however, that it can only be applied when θ is low-dimensional; as such, it is useful for our problem only in the $I(1)$ and local-to-unity stochastic trend model for $r = k = 1$.

On the other hand, we discuss how to numerically determine a reasonably low upper bound for cases where θ is of high dimension, to be compared to the power of the *JW* statistic in Section 6 below. This latter approach exploits both the multivariate normal nature of the inference problem, and the existence of an *ad hoc* test that is known to control size, and is thus more specific to the cointegration testing problem.

5.4.1 Low Dimensional Nuisance Parameter

Suppose that $\text{LR}_\Lambda = h(U)/\int f_\theta(U)d\Lambda(\theta)$ is a continuous random variable for any Λ, so that by the Neyman-Pearson Lemma, φ_Λ is of the form $\varphi_\Lambda = \mathbf{1}[\text{LR}_\Lambda > \text{cv}_\Lambda]$, and cv_Λ is chosen so that φ_Λ is level α under H_Λ, i.e. $\alpha = \int \int \varphi_\Lambda f_\theta d\mu d\Lambda(\theta)$. Then by Lemma 2, the power of φ_Λ, $\beta_\Lambda = \int \varphi_\Lambda h d\mu$, is an upper bound on the power of any test that is level α under H_0. If Λ is not the least favorable distribution, then φ_Λ is not of size α under H_0, i.e. $\sup_{\theta \in \Theta} \int \varphi_\Lambda f_\theta d\mu > \alpha$. Now consider a version of φ_Λ with a size corrected critical value $\text{cv}_\Lambda^c > \text{cv}_\Lambda$, that is $\varphi_\Lambda^c = \mathbf{1}[\text{LR}_\Lambda > \text{cv}_\Lambda^c]$ with $\sup_{\theta \in \Theta} \int \varphi_\Lambda^c f_\theta d\mu = \alpha$. The size adjusted test φ_Λ^c is of level α under H_0 by construction, so *if* its power $\beta_\Lambda^c = \int \varphi_\Lambda^c h d\mu$ is only marginally less than β_Λ, then it is an almost efficient test, since β_Λ is an upper power bound for all level α tests. Also, by definition, the least upper bound on the power of level α tests must be sandwiched between β_Λ^c and β_Λ, so that we would also have identified a good approximation to the least upper bound.

The challenge is thus to find an appropriate Λ. This is difficult because, in general, no closed form solutions are available for the size and power of tests, so that these must be approximated by Monte Carlo integration. Brute force searches for an appropriate Λ are thus not computationally feasible. The idea is to exploit numerical advantages of discrete distributions for Λ, that have point masses at only N points, and to smooth out the Monte Carlo integration estimates of size and power, so that gradient methods can be employed. The suggested algorithm is related to, but distinct from those developed in Nelson (1966), Kempthorne (1987) and Sriananthakumar and King (2006), and is described in detail in the appendix.

5.4.2 High Dimensional Nuisance Parameter

The dimension of θ can be very large in our problem: even when $r = k = 1$, the model with unrestricted stochastic trend leads to θ of dimension $q^2 + q(q+1)/2$ so that θ contains 222 elements when $q = 12$. Approximating the least upper power bound directly then becomes a numerically intractable problem. This motivates a computationally practical method for computing a *low* (as oppose to *least)* upper power bound.

The method restricts Λ so that it is degenerate with all mass on a single point, say θ^*, which is chosen so that the null distribution of the maximal invariant of Theorem 2 is close to its distribution under the alternative. Since the density of the maximal invariant is quite involved, θ^* is usefully approximated by a choice that makes the multivariate normal distribution of $\text{vec}(Y, X)$ under the null close to its distribution under the alternative, as measured by a convenient metric. To be specific, let Σ_1 denote the covariance matrix of $\text{vec}(Y, X)$ under a specific $I(1)$ alternative as described in Subsection 5.2.1 above (that is, for specific values of $B = B_1$ and $R = R_1$), let $\Sigma_0(\theta)$ with $\theta \in \Theta$ be the covariance matrix of $\text{vec}(Y, X)$ under the null for the relevant restrictions on the stochastic trend, and define the $nq \times nq$ matrix (recall that $n = r + q$)

$$
A(\gamma) = \left[\begin{array}{cc} \gamma_{yz} \otimes I_q & 0 \\ \gamma_{xz} \otimes I_q & \gamma_{xv} \otimes I_q \end{array} \right]
$$

where γ_{yz} is $r \times r$, γ_{xz} is $k \times r$, and γ_{xv} is $k \times k$. This yields $A(\gamma) \text{vec}(Y, X) \sim \mathcal{N}(0, A(\gamma)\Sigma_0(\theta)A(\gamma)')$. Denote the Kullback-Leibler divergence between the $nq \times 1$ distributions $\mathcal{N}(0, \Sigma_1)$ and $\mathcal{N}(0, \Sigma_0)$ as $K(\Sigma_1, \Sigma_0) = \frac{1}{2}\ln(\det \Sigma_1 / \det \Sigma_0) + \frac{1}{2}\text{tr}(\Sigma_0^{-1}\Sigma_1) - nq$. The value of θ^* is chosen to numerically solve

$$
\min_{\gamma \in \mathbb{R}^{r^2+k^2+kr}} K(\Sigma_1, A(\gamma)\Sigma_0(\theta^*)A(\gamma)') = \min_{\theta \in \Theta, \gamma \in \mathbb{R}^{r^2+k^2+kr}} K(\Sigma_1, A(\gamma)\Sigma_0(\theta)A(\gamma)'), \qquad (28)
$$

that is, θ^* numerically minimizes the Kullback-Leibler divergence (or KLIC) between the null and alternative densities of (Y, X), allowing for transformations as described by $A(\gamma)$ under the null. While these transformations do not affect the implied distribution of the maximal invariant, they do in general lead to a different θ^*, which we found to yield a slightly lower upper bound. The minimization problem is over a high dimensional parameter, but the objective function is quickly computed and well behaved, so that numerical minimization

is feasible. (And of course, the validity of the power bound from Lemma 2 based on the (degenerate) distribution Λ that puts all mass on the numerical minimizer does not require that the numerical minimizer coincides with the global minimizer. Said differently, numerical error in the solution of (28) will not invalidate the resulting bound.)

The power of the Neyman-Pearson test of $H_\Lambda : \text{vec}(Y, X) \sim \mathcal{N}(0, \Sigma_0(\theta^*))$ against $H_1 : \text{vec}(Y, X) \sim \mathcal{N}(0, \Sigma_1)$ based on the maximal invariant of Theorem 2 is not necessarily a least upper power bound for invariant tests of $H_0 : \text{vec}(Y, X) \sim \mathcal{N}(0, \Sigma_0(\theta))$, $\theta \in \Theta$, against H_1. Lemma 2, however, implies that it is an upper power bound, and the numerical results in the next section suggest that it is reasonably close to the least upper power bound in models with low-dimensional nuisance parameter θ, where the least upper bound can be directly approximated. Moreover, in many other cases it is close to the power of the JW test, so again it is close the least upper power bound, as the least upper bound cannot be below the power of the JW test. Thus, the power bound associated with minimizing the Kullback-Leibler divergence (28) proves to be a practical and useful approximation to the least upper power bound in this problem.

Note that the upper bound on power that results from this approach follows only from the constraint that tests need to control size for the particular value $\theta = \theta^*$. The range of the models allowed under a particular stochastic trend model, as described by the size of the set Θ, does not play any role. Thus, even if one were to weaken the robustness constraint of tests to control size over a much smaller set $\Theta_s \subset \Theta$, one would still obtain the same bound as long as $\theta^* \in \Theta_s$. The crucial question for the appeal of the robustness constraint is thus whether it makes sense to insist on size control for the particular stochastic trend model that leads to $\text{vec}(Y, X) \sim \mathcal{N}(0, \Sigma_0(\theta^*))$.

A natural question to ask is whether, as an empirical matter, a realization of the trend from the $\Sigma_0(\theta^*)$ model looks, in some sense, reasonable. In this regard, it useful to remark that θ^* is chosen precisely to minimize KLIC relative to the standard $I(1)$ model under the alternative. Ignoring the transformation induced by $A(\gamma)$, since $\text{vec}\, Y \sim \mathcal{N}(0, I_{rq})$ for any value of θ under the null hypothesis, the KLIC criterion can usefully be thought of as attempting to approximate the conditional distribution of X given Y in the alternative model. In fact, in the unrestricted model, Lemma 1 implies that one can precisely replicate this conditional distribution under the null hypothesis.[8] Draws from $\text{vec}(Y, X) \sim \mathcal{N}(0, \Sigma_0(\theta^*))$

---

[8]Without the invariance restriction (21), this observation would lead to an analytic least favorable dis-

are thus approximately rationalizable as draws from the model with standard $I(1)$ stochastic trends, except for the difference in the marginal distribution of Y.

5.5 Power Bounds for $r = 1$ and $k > 2$

In the next section we will calculate power bounds for selected values of r and k including $r = 1$ and $k = 2$. In this subsection we show that the resulting bounds also serve as power bounds for models with $r = 1$ and all values of $k > 2$.

To see why, first consider the alternative $I(1)$ model as described in subsection 5.2.1, $Y = Z + VB'$ and $X = V$. Let P be a $k \times k$ orthonormal matrix whose last $k - 2$ rows are orthogonal to R and B, and whose second row is orthogonal to R. Partition $X = (X_{12}, X_{3k})$, where X_{12} contains the first two columns of X and X_{3k} contains the remaining $k - 2$ columns. By invariance, there is no loss in generality in setting $X = \tilde{X}P = (\tilde{X}_{12}, \tilde{X}_{3k})P$, so that $Y = Z + \tilde{X}PB' = Z + \tilde{X}_{12}B'_{12}$, where $\tilde{X}_{12}$ and B_{12} are the first two columns of $\tilde{X}$ and B, respectively, and the last $k - 1$ columns of $\tilde{X}$ (and thus $\tilde{X}_{3k}$) are independent of Z. The group of transformations

$$(Y, \tilde{X}_{12}, \tilde{X}_{3k}) \rightarrow (YA_{yy}, \tilde{X}_{12}\tilde{A}_{xx} + YA_{xy}, \tilde{X}_{3k}) \tag{29}$$

for nonsingular A_{yy} and $\tilde{A}_{xx}$ is a subset of the transformations $(Y, \tilde{X}) \rightarrow (YA_{yy}, \tilde{X}A_{xx} + YA_{xy})$, so the best invariant test to (29) is as least as powerful as the best invariant test to (21). Let $\tilde{Q}_{12}$ be a maximal invariant to $(Y, \tilde{X}_{12}) \rightarrow (YA_{yy}, \tilde{X}_{12}\tilde{A}_{xx} + YA_{xy})$, such that $\{\tilde{Q}_{12}, \tilde{X}_{3k}\}$ is a maximal invariant to (29). Since $\tilde{X}_{3k}$ is independent of $(Y, \tilde{X}_{12})$, the density of $\{\tilde{Q}_{12}, \tilde{X}_{3k}\}$ under the alternative factors as $f_{a,\tilde{Q}_{12}} \cdot f_{a,\tilde{X}_{3k}}$.

For all null models discussed in subsection 5.2.2, it is possible to choose $X = (X_{12}, X_{3k}) = V$ in a way such that X_{3k} is independent of X_{12} with marginal distribution $f_{0,X_{3k}} = f_{a,\tilde{X}_{3k}}$, (i.e. it corresponds to the $I(1)$ model) and the possibilities for X_{12} and its relationship with Z are the same as in the version of the model with $k = 2$. Thus, with this choice, the term $f_{a,\tilde{X}_{3k}}$ cancels in the likelihood ratio test of the maximal invariant to (29), and the testing

---

tribution result: Factor the density of (Y, X) into a the product of the density of Y, and the density of X given Y. Since the latter cancels with the appropriate choice of θ under the conditions of Lemma 1, this shows that the Neyman-Pearson test is a function of Y only.

problem corresponds precisely to the model with $k = 2$.[9] An upper bound for the model with $r = 1$ and $k = 2$ is therefore also an upper bound for the model with $r = 1$ and $k > 2$.

5.6   Deterministic trends

Thus far, the analysis has ignored deterministic linear trends, and for the sake of brevity we will not offer a detailed treatment of tests that incorporate them. That said, it is useful to offer a brief outline. In this regard, it is useful to distinguish between the usual notion of cointegration in which z_t is $I(0)$ and what has been called "stochastic cointegration" by Ogaki (1988) and Ogaki and Park (1998) in which z_t may contain a linear trend in addition to an $I(0)$ component. In this latter case, it is reasonable to consider tests that are invariant to adding linear trends to the data, and Müller and Watson (2007) show how this additional invariance can be handled by an alternative choice of Ψ functions used to compute the low-frequency transformations of the data. Thus, incorporating stochastic cointegration is quite straightforward.

Perhaps the more relevant notion of cointegration maintains the assumption that z_t is $I(0)$ under the null, but allows v_t to contain a linear trend. This introduces two considerations: First, under the null hypothesis, Y-only tests remain unaffected, but the additional constraint that tests have to control size for any value of the linear trend coefficient can only further reduce the upper bounds on tests involving (Y, X). Second, it may desirable to specify the alternative so that the test maximizes power for stochastic trend processes that include linear trends. With a Gaussian weighting function over the trend coefficient, maximizing weighted average power becomes equivalent to maximizing power against different Σ_1 values, and associated power bounds could be computed using the methods outlined above.

---

[9]This is not strictly true for the stationary G-model, which excludes $I(1)$ stochastic trends. But the low-frequency transformation of the suitably scaled stationary local-to-unity model converges in mean squared to the $I(1)$ model as the local-to-unity parameter approaches zero (cf. Section 2.4 of Müller and Watson (2007)), so that the additional discriminatory power from X_{3k} can be made arbitrarily small, and the conclusion continues to hold.

6 Results

This section provides numerical results for bounds on the power of 5% level tests based on (Y, X) against the $I(1)$ alternative that control size for various restrictions on the stochastic trend model, as discussed in Section 5. These bounds are compared to power of Y-only tests against this alternative, as derived in Section 4.

Table 2 summarizes power bounds for various models with $q = 12$, and with $r = 1$ and $k = 1$ (panel A), $r = 1$ and $k = 2$ (panel B), and $r = 2$ and $k = 1$ (Panel C). Numerical results for larger values of $n = r + k$ are not reported because of the large number of calculations required to evaluate the density in large models.

Power depends on the values of B and R under the alternative, and results are presented for various values of these parameters. Because of invariance, when $r = 1$ (as in panels A and B), or $k = 1$ (as in panel C), the distribution of the maximal invariant depends on B and R only through $||B||$, $||R||$, and, if $||R|| > 0$, on $\text{tr}(R'B)/(||B|| \cdot ||R||)$. Thus, in panel A, where $r = k = 1$, results are shown for two values of $|B|$, three values of $|R|$ and for $R \cdot B < 0$ and $R \cdot B > 0$, while panels B and C show results for three values of $\text{tr}(R'B)/(||B|| \cdot ||R||)$ when $||R|| > 0$. As discussed in subsection 5.5 above, the power bounds in panel B for $r = 1$ and $k = 2$ also carry over to models with $r = 1$ and arbitrary $k > 2$.

Looking first at panel A, the first four columns of the table summarize the power bounds computed using θ^*, the KLIC minimizing value for θ as described in subsection 5.4.2. Only four columns are needed to summarize all of the models because the G model and the diagonal G model are identical when $k = 1$, and Lemma 1 shows the G-model imposes no additional restrictions on $\Sigma_{(Y,X)}$ when $r \leq k$. The final two columns of the table show the results for a direct numerical approximation to the least upper power bound, based on the algorithm discussed in subsection 5.4.1. This algorithm can only be implemented when the number of nuisance parameters is small, and results are shown for the local-to-unity model (where C and R are the only scalar nuisance parameters) and the $I(1)$ models (where R is the only nuisance parameter). As discussed in the appendix, the upper bounds are constructed so that they are at most 2.5 percentage points above the actual least upper bounds (ignoring Monte Carlo integration error), and thus roughly correspond to the power of optimal tests for testing H_0 versus H_1.

Looking across the entries in panel A, three results are evident. First, and not sur-

prisingly, restricting tests so that they control size for the unrestricted trend process has a non-negligible power cost. For example, when $|B| = 7$, and $R = 0$, the power bound is 0.36, for tests that control size for unrestricted trends, the bound increases to 0.41 for tests that control size for stationary trends, and increases to 0.50 for tests that control size for local-to-unity or $I(1)$ trend processes. Second, as noted by Stock and Watson (1996), Jansson and Moreira (2006), and others, the local-to-unity model yields asymmetric power functions when $R \neq 0$. For example in the local-to-unity model with $|R| = 0.9$ and $|B| = 7$, the power bound is 0.67 for alternatives with $R \cdot B > 0$, but increases to 0.96 when $R \cdot B < 0$. This asymmetry is also present in the stationary model, but not in the $I(1)$ or general trend model. Third, the power bounds associated with the KLIC minimizing value for θ are only slightly larger than the (approximate) least upper power bounds computed for the $I(1)$ and local-to-unity models, at least for $|B| = 7$. Evidently then, in these cases the KLIC minimizers provide a good approximation to the least upper power bound.

Panel B shows results when $r = 1$ and $k > 1$. This panel does not include entries for approximate least upper power bounds, because of the computational burden associated with the additional nuisance parameters. Because $k > 1$, the diagonal-G trend imposes restrictions on $\Sigma_{(Y,X)}$ under the null, and results are presented for tests that control size under this restriction. The power bounds for the restricted models shown in panel B are lower than the bounds shown in panel A because, for example, the tests must control size for the nuisance parameters in the $k \times 1$ correlation vector R, and k is larger in panel B than in panel A.

The most important conclusion to be drawn form panels A and B is that, for all practical purposes, the JW test introduced in Section 4 achieves the power bound of the best invariant test that controls size for the unrestricted trend process. Power results are reported for the $JW(||B||)$ statistic, which yields the power envelope for Y-only tests, and for $JW(10/\sqrt{r})$, the feasible point-optimal test with critical values tabulated in Table 1. From the results in panels A and B, the power of these two tests are very close, which indicates that there is only a small loss in power associated with not knowing B under the alternative. Because the JW test essentially achieves the power bound associated with tests that control size for unrestricted trends for $r = 1$ and any value of k, the test is approximately optimal test in this case.[10]

_____

[10]Recall from the discussion in Section 4 that the power of the JW test depends on B only through $||B||$

While computational limits prevent us for showing power bounds with $r > 1$ and arbitrary k, panel C shows results for the model with $r = 2$ and $k = 1$. Because $r > k$, this model has two important features not present in panels A and B. First, the null covariance matrix $\Sigma_{(Y,X)}$ is restricted when $H(s,t) = G(s,t)S_v$, so that this restricted trend specification is added to the table. (Because $k = 1$, $G(s,t)$ is diagonal, and this restriction is dropped from the table.) Second, the efficient Y-only test is not the JW test. Rather, as was evident in Figure 1, the efficient Y-only test exploits the restriction that $k < r$, and, from the expression for Σ_{YY} given in (25), the test now depends on R. The final column of panel C shows the power-envelope for the best Y-only test. In contrast to results with $r = 1$, there is now some power gain associated with exploiting the X data, even for tests that control size for the unrestricted trend model. That is, the power bound shown in the first column of the table (for the (X, Y)-test with an unrestricted trend under the null) is higher than the power-envelope for the Y-only test shown in the final column. The difference in power is small when $||R||$ is small, but is non-negligible when $||R||$ is large.

As a by-product of the direct approximations of the least upper power bound in the low-dimensional models for $r = k = 1$, the algorithm described in subsection 5.4.1 also yields a 5% level test (that is, size is at most 5%) whose power is within 2.5 percentage points of the bound. These tests could thus in principle be employed for inference in the bivariate model with local-to-unity local or $I(1)$ stochastic trends, and they (approximately) maximize power against the $I(1)$ point alternative for a specific value of R (and B). Since R is unknown in practice, this is of limited appeal. But the algorithm can also be employed to determine tests that approximately maximize weighted average power against a set of alternatives in these models. Table 3 displays the power of 5% level tests whose equally weighted average power against alternatives with $B = 7$ and the 20 values of $R = -0.95, -0.85, \cdots, 0.95$ is within 2.5 percentage points of the power bound on that weighted average power, along with the bounds on the power envelope from Table 2. The approximately weighted average power maximizing test statistics are of the form $\sum_{j=1}^{20} \frac{1}{20} f_1(R_j, 7) / \sum_{j=1}^{N} p_j f_0(\theta_j)$, where $f_1(R, B)$ is the density of the maximal invariant of Theorem 2 under the local alternative in the $I(1)$ model, f_0 is the density under the null model for a particular value of the nuisance parameter θ describing R and, in local-to-unity model, C, and N, θ_j and p_j are numerically

when $r = 1$, for any value of k.

determined by the algorithm.[11] It is apparent that there is some cost to not knowing R under the alternative, as the power of the (approximately) weighted average power maximizing test falls up to 11 percentage points short of the power envelope for both stochastic trend models, even against the alternative with $B = 7$.

We performed the same set of computations shown in Tables 2 and 3 also for $q = 6, 9, 15$ and 24, but do not report detailed results for brevity. All qualitative conclusions continue to hold: The JW test almost achieves the power bound in the unrestricted model for $r = 1$ and any value of k, there are substantial gains in power for highly restricted stochastic trend models, and the power bounds obtained through minimizing KLIC are relatively close to those obtained by directly approximating the least upper bound.

7   Concluding Remarks

This paper studies inference about the cointegrating vector in a framework in which the common stochastic trends, and their interaction with the error correction terms, is modelled in a flexible way beyond the standard $I(1)$ framework. The problem is studied with the low-frequency transformation approach suggested by Müller and Watson (2007). The paper derives bounds on the power of tests that control size over flexible stochastic trend specifications, and which maximize power against alternatives with the usual $I(1)$ trend. We find that a low-frequency version of Wright's (2000) test (JW) essentially achieves the upper power bound in the model with $r = 1$ cointegrating vectors.

The implication for applied work is that, at least in the model with $r = 1$, approximately efficient robust inference about the cointegrating vector may be carried out using this test. The test is simple to compute and is robust in two ways. First, it is robust to arbitrary autocorrelation properties in the error correction term above the pre-specified low-frequency band. Second, it is robust to the precise nature of persistence, as its rejection probability under the null hypothesis does not depend on the nature of the stochastic trend. As in Wright (2000), confidence sets for the cointegrating vector can easily be obtained be inverting the test.

---

[11]For instance, with the $I(1)$ stochastic trend under the null hypothesis, $\theta = R$, $N = 4$, $(p_1, \theta_1) = (0.8170, 0.35)$, $(p_2, \theta_2) = (0.1596, 0.75)$, $(p_3, \theta_3) = (0.0228, 0.85)$ and $(p_4, \theta_4) = (0.0006, 0.97)$.

This paper also provides a partial solution to the question of efficient inference about the value of cointegrating vectors if strong *a priori* knowledge about the stochastic trend is available: with a tightly parametrized model for the stochastic trend, we show how an approximately (weighted average) power maximizing test can be determined numerically. The suggested method is generic in the sense that it computes an approximately efficient test in the presence of a low dimensional nuisance parameters under the null hypothesis. This type of problem arises naturally in nonstandard testing problems, so we would expect the method to be useful also in other contexts.

The construction of efficient tests for the value of the cointegrating vector that control size for an unrestricted trend model when $r > 1$ remains an open question. However, the power bounds computed here provide a useful check for the efficiency of *ad hoc* tests that might be suggested for this problem.

As we have shown, the JW test for the value of the cointegrating vectors is approximately efficient against alternatives where the system under consideration is cointegrated, but the hypothesized value of the cointegrating vector is wrong. A different kind of alternative arises from a system that is not cointegrated. A local deviation from cointegration is naturally modelled by a putative error correction term that is the sum of an $I(0)$ component and a small trend component. Power bounds for these alternatives can also be studied using the methods developed in this paper, although the details will depend on the structure of the additional stochastic trends that are present under the alternative.

A Appendix

A.1 Proof of Theorem 1

The proof to part (a) mimics the proof to Theorem 2 and is omitted. To prove part (b), note that because of invariance, we can set $\tilde{V}_{YY} = I_r$ without loss of generality, so that $\det \Sigma_{YY} = (\det \tilde{\Sigma}_{YY})^r$, $\Omega_Y = (I_r \otimes Y'\tilde{\Sigma}_{YY}^{-1}Y)$ and $(\det \Omega_Y)^{-1/2} = \det(Y'\tilde{\Sigma}_{YY}^{-1}Y)^{-r/2}$. Since $(\operatorname{vec}\omega_Y)'\Omega_Y(\operatorname{vec}\omega_Y) = \operatorname{tr}(\omega_Y' Y'\tilde{\Sigma}_{YY}^{-1}Y\omega_Y)$, the density in part (a) of the Theorem becomes proportional to

$$(\det \tilde{\Sigma}_{YY})^{-r/2} \det(Y'\tilde{\Sigma}_{YY}^{-1}Y)^{-r/2} \int |\det \omega_Y|^{q-r} \exp[-\tfrac{1}{2}\operatorname{tr}(\omega_Y'Y'\tilde{\Sigma}_{YY}^{-1}Y\omega_Y)]d(\operatorname{vec}\omega_Y).$$

Let $\tilde{\omega}_Y = (Y'\tilde{\Sigma}_{YY}^{-1}Y)^{1/2}\omega_Y$, so that $|\det \omega_Y|^{q-r} = \det(Y'\tilde{\Sigma}_{YY}^{-1}Y)^{-(q-r)/2}|\det \tilde{\omega}_Y|^{q-r}$ and $\operatorname{vec}\omega_Y = (I_r \otimes (Y'\tilde{\Sigma}_{YY}^{-1}Y)^{-1/2})\operatorname{vec}\tilde{\omega}_Y$, and the Jacobian determinant of the transformation from ω_Y to $\tilde{\omega}_Y$ is $\det(I_r \otimes (Y'\tilde{\Sigma}_{YY}^{-1}Y)^{-1/2}) = (Y'\tilde{\Sigma}_{YY}^{-1}Y)^{-r/2}$. Thus, the density is proportional to

$$(\det \tilde{\Sigma}_{YY})^{-r/2} \det(Y'\tilde{\Sigma}_{YY}^{-1}Y)^{-q/2} \int |\det \tilde{\omega}_Y|^{q-r} \exp[-\tfrac{1}{2}\operatorname{tr}(\tilde{\omega}_Y'\tilde{\omega}_Y)]d(\operatorname{vec}\tilde{\omega}_Y),$$

and the result follows.

A.2 Proof of Theorem 2

Write $Y = (Y_1', Y_2', Y_3')'$ and $X = (X_1', X_2', X_3')'$, where Y_1 and X_1 have r rows, and Y_2 and X_2 have k rows. Consider the one-to-one mapping $h : \mathbb{R}^{q \times n} \mapsto \mathbb{R}^{q \times n}$ with

$$h(Y, X) = Q = \begin{pmatrix} Q_{Y1} & Q_{X1} \\ Q_{Y2} & Q_{X2} \\ Q_{Y3} & Q_{X3} \end{pmatrix} = \begin{pmatrix} Y_1 & Y_1^{-1}X_1 \\ Y_2(Y_1)^{-1} & X_2 - Y_2Y_1^{-1}X_1 \\ Y_3(Y_1)^{-1} & (X_3 - Y_3Y_1^{-1}X_1)(X_2 - Y_2Y_1^{-1}X_1)^{-1} \end{pmatrix}.$$

A straightforward calculation shows that $(\operatorname{vec}Q_{Y2}', \operatorname{vec}Q_{Y3}', \operatorname{vec}Q_{X3}')$ is a maximal invariant to (21). The inverse of h is given by

$$h^{-1}(Q) = \begin{pmatrix} Q_{Y1} & Q_{Y1}Q_{X1} \\ Q_{Y2}Q_{Y1} & Q_{X2} + Q_{Y2}Q_{Y1}Q_{X1} \\ Q_{Y3}Q_{Y1} & Q_{X3}Q_{X2} + Q_{Y3}Q_{Y1}Q_{X1} \end{pmatrix}.$$

33

Using matrix differentials (cf. Chapter 9 of Magnus and Neudecker (1988.)), a calculation shows that the Jacobian determinant of h^{-1} is equal to $(\det Q_{Y1})^{q-r+k}(\det Q_{X2})^{q-k-r}$. The density of Q is thus given by

$$(2\pi)^{-qn/2}(\det \Sigma_{(Y,X)})^{-1/2}|\det Q_{Y1}|^{q-r+k}|\det Q_{X2}|^{q-k-r}\exp[-\tfrac{1}{2}(\operatorname{vec} h^{-1}(Q))'\Sigma_{(Y,X)}^{-1}(\operatorname{vec} h^{-1}(Q))]$$

and we are left to integrate out Q_{Y1}, Q_{X1} and Q_{X2} to determine the density of the maximal invariant.

Now consider the change of variables from Q_{Y1}, Q_{X1}, Q_{X2} to ω_Y, ω_X and ω_{YX}

$$
\begin{aligned}
Q_{Y1} &= Y_1\omega_Y \\
Q_{X1} &= \omega_Y^{-1}Y_1^{-1}X_1\omega_X - \omega_Y^{-1}\omega_{YX} \\
Q_{X2} &= (X_2 - Y_2Y_1^{-1}X_1)\omega_X
\end{aligned}
$$

with Jacobian determinant $(\det Y_1)^r(\det(X_2 - Y_2Y_1^{-1}X_1))^k \det(-\omega_Y)^{-k}$. Noting that with this change, $h^{-1}(Q) = (Y\omega_Y, X\omega_X - Y\omega_{YX})$, we find that the density of the maximal invariant is equal to

$$\int (2\pi)^{-qn/2}(\det \Sigma_{(Y,X)})^{-1/2}|\det Y_1|^{q+k}|\det(X_2 - Y_2Y_1^{-1}X_1)|^{q-r}|\det \omega_Y|^{q-r}|\det \omega_X|^{q-k-r}$$

$$\cdot\exp[-\tfrac{1}{2}\operatorname{vec}(Y\omega_Y, X\omega_X - Y\omega_{YX})'\Sigma_{(Y,X)}^{-1}\operatorname{vec}(Y\omega_Y, X\omega_X - Y\omega_{YX})]d(\operatorname{vec}\omega_Y', \operatorname{vec}\omega_X', \operatorname{vec}\omega_{YX}')'.$$

Since $\operatorname{vec}(Y\omega_Y, X\omega_X - Y\omega_{YX}) = D_{YX}\operatorname{vec}(\omega_Y, \omega_X) - V_{0Y}\operatorname{vec}(\omega_{YX})$, we have

$$
\begin{aligned}
\operatorname{vec}(Y\omega_Y, &X\omega_X - Y\omega_{YX})'\Sigma_{(Y,X)}^{-1}\operatorname{vec}(Y\omega_Y, X\omega_X - Y\omega_{YX}) \\
&= \operatorname{vec}(\omega_Y, \omega_X)'D_{YX}'\Sigma_{(Y,X)}^{-1}D_{YX}\operatorname{vec}(\omega_Y, \omega_X) \\
&\quad - 2\operatorname{vec}(\omega_Y, \omega_X)D_{YX}'\Sigma_{(Y,X)}^{-1}V_{0Y}\operatorname{vec}(\omega_{YX}) + \operatorname{vec}(\omega_{YX})'V_{0Y}'\Sigma^{-1}V_{0Y}\operatorname{vec}(\omega_{YX}).
\end{aligned}
$$

The result now follows from integrating out ω_{YX} by 'completing the square'.

A.3 Proof of Lemma 1

We first establish a preliminary result.

Lemma 3 *For any $t > 0$ and integer κ, the functions $\Psi_l : [0,t] \mapsto \mathbb{R}$ with $\Psi_l(s) = \sqrt{2}\cos(\pi ls)$, $l = 1, \cdots, \kappa$ are linearly independent.*

Proof. Choose any real constants c_j, $j = 1, \cdots, \kappa$, so that $\sum_{j=1}^{\kappa} c_j \Psi_j(s) = 0$ for all $s \in [0, t]$. Then also $\sum_{j=1}^{\kappa} c_j \Psi_j^{(i)}(0) = 0$ for all $i > 0$, where $\Psi_j^{(i)}(0)$ is the ith (right) derivative of Ψ_j at $s = 0$. A direct calculation shows $\Psi_j^{(i)}(0) = (-1)^{i/2} \sqrt{2} (\pi j)^i$ for even i. It is not hard to see that the $\kappa \times \kappa$ matrix with j, ith element $(-1)^{i/2} (\pi j)^i$ is nonsingular, so that $\sum_{j=1}^{\kappa} c_j \Psi_j^{(i)}(0) = 0$ for $i = 2, 4, \cdots, 2\kappa$ can only hold for $c_j = 0$, $j = 1, \cdots, \kappa$. ∎

For the proof Lemma 1 we construct $H(s, t)$ such $\text{vec}\, Z = \int_0^1 (I_r \otimes \Psi(t)) S_z dW(t)$ and $\text{vec}\, V = \int_0^1 \int_t^1 (H(s, t) \otimes \Psi(s)) ds\ dW(t)$ have the specified covariance matrix. The proof of the slightly more difficult part (b), where $H(s, t) = G(s, t) S_v$, is based on the following observations:

(i) Ignoring the restriction on the form of $\text{vec}\, V$, it is straightforward to construct an appropriate multivariate normal vector $\text{vec}\, V$ from a linear combination of $\text{vec}\, Z$ and ζ, where $\zeta \sim \mathcal{N}(0, I_{kq \times kq})$ independent of Z.

(ii) Suppose that $R = S$ was allowed, where $S = (I_r, 0_{r \times (k-r)})$. Then $S_z = SS_v$, $\text{vec}\, Z \sim \int_0^1 F_Z(t) S_v dW(t)$ for $F_Z(t) = S \otimes \Psi(t)$, and one can also easily construct ζ as in (i) via $\zeta = \int_0^1 F_\zeta(t) S_v dW(t)$ by an appropriate choice of F_ζ. Since Ito-Integrals are linear, one could thus write $\text{vec}\, V = \int_0^1 F(t) S_v dW(t)$ with F a linear combination of F_Z and F_ζ, using observation (i).

(iii) For any matrix function $F : [0, 1] \mapsto \mathbb{R}^{kq \times k}$ that is equal to zero on the interval $(1 - \varepsilon, 1]$ for some $\varepsilon > 0$, one can set $G(s, t) = (I_k \otimes \Psi(s)' J(t)^{-1}) F(t)$, where $J(t) = \int_t^1 \Psi(s) \Psi(s)' ds$ and obtain $\int_0^1 \int_t^1 (G(s, t) \otimes \Psi(s)) ds\ S_v dW(t) = \int_0^1 F(t) S_v dW(t)$, since for any matrix A with k rows and vector v, $A \otimes v = (I_k \otimes v) A$.

The following proof follows this outline, but three complications are addressed: $R = S$ is not allowed; the matrix function F needs to be zero on the interval $(1 - \varepsilon, 1]$, which does not happen automatically in the construction in (ii); one must verify that the process $\int_0^s G(s, t) S_v dW(t)$ admits a cadlag version.

Set S_z to be the first r rows of I_n. Since $\Psi_l(1 - s) = (-1)^l \Psi_l(s)$ for all $l \geq 1$, Lemma 3 implies that $J(t) = \int_t^1 \Psi(s) \Psi(s)' ds$ and $I_q - J(t)$ are nonsingular for any $t \in (0, 1)$. The $rq \times 1$ random vector $\text{vec}\, Z_\varepsilon = \int_0^{1-\varepsilon} (S_z \otimes \Psi(s)) dW(s)$ thus has nonsingular covariance matrix

$I_r \otimes \Sigma_q^\varepsilon$, where $\Sigma_q^\varepsilon = I_q - J(1 - \varepsilon)$. Also, since

$$\Sigma^* = \begin{pmatrix} I_r \otimes I_q & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

is positive definite, so is $I_{rq} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, so that we can choose $0 < \varepsilon < 1$ such that $I_r \otimes \Sigma_q^\varepsilon - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is positive definite. With that choice of ε, also $\Sigma_{22} - \Sigma_{21}(I_r \otimes \Sigma_q^\varepsilon)^{-1}\Sigma_{12}$ is positive definite.

For part (a) of the lemma, define the $[0, 1] \mapsto \mathbb{R}^{kq \times n}$ function $F_a(t) = A_a(I_n \otimes \Psi(t))$, where $A_a = (A_{a1}, A_{a2})$ with $A_{a1} = \Sigma_{21}(I_r \otimes \Sigma_q^\varepsilon)^{-1}$ and $A_{a2} = (\Sigma_{22} - \Sigma_{21}(I_r \otimes \Sigma_q^\varepsilon)^{-1}\Sigma_{12})^{1/2}(I_k \otimes (\Sigma_q^\varepsilon)^{-1/2})$.

For part (b) of the lemma, choose $0 < \rho < 1$ so that $\Sigma_{22} - \rho^{-2}\Sigma_{21}(I_r \otimes \Sigma_q^\varepsilon)^{-1}\Sigma_{12}$ is positive definite. Choose S_v to be the first k rows of I_n multiplied by ρ, so that $R = S_z S_v' = \rho S$. Let $\tilde{\Psi}_1(s)$ be scaled residuals of a continuous time projection of $\mathbf{1}[s \le 1 - \varepsilon]\Psi_{q+1}(s)$ on $\{\mathbf{1}[s \le 1 - \varepsilon]\Psi_l(s)\}_{l=1}^q$ on the unit interval, and let $\tilde{\Psi}_j(s)$, $j = 2, \cdots, kq$ be the scaled residuals of continuous time projection of $\mathbf{1}[s \le 1 - \varepsilon]\Psi_{q+j}(s)$ on $\{\mathbf{1}[s \le 1 - \varepsilon]\Psi_l(s)\}_{l=1}^q$ and $\{\mathbf{1}[s \le 1 - \varepsilon]\tilde{\Psi}_l(s)\}_{l=1}^{j-1}$. By Lemma 3, $\tilde{\Psi}_j(s)$, $j = 1, \cdots, kq$, are not identically zero, and we can choose their scale to make them orthonormal. Let $\tilde{\Psi}(s) = (\tilde{\Psi}_1(s), \cdots, \tilde{\Psi}_{kq}(s))'$, the $k \times 1$ vector $\iota_k = (1, 0, \cdots, 0)'$, and $A_b = (A_{b1}, A_{b2})$ with $A_{b1} = \rho^{-1}\Sigma_{21}(I_r \otimes \Sigma_q^\varepsilon)^{-1}$ and $A_{b2} = (\Sigma_{22} - \rho^{-2}\Sigma_{21}(I_r \otimes \Sigma_q^\varepsilon)^{-1}\Sigma_{12})^{1/2}$. Now define the $[0, 1] \mapsto \mathbb{R}^{kq \times n}$ function

$$F_b(t) = A_b \begin{pmatrix} S \otimes \Psi(t) \\ \iota_k' \otimes \tilde{\Psi}(t) \end{pmatrix} S_v.$$

For both parts, that is for $i \in \{a, b\}$, set

$$H_i(s, t) = (I_k \otimes \Psi(s)' J(t)^{-1})F_i(t) \quad \text{for } t \in [0, 1 - \varepsilon]$$

and $H_i(s, t) = 0$ otherwise. With this choice

$$
\begin{aligned}
\text{vec } V_i &= \int_0^1 \int_t^1 (H_i(s, t) \otimes \Psi(s))ds \; dW(t) \\
&= \int_0^{1-\varepsilon} \int_t^1 ((I_k \otimes \Psi(s)\Psi(s)' J(t)^{-1})F_i(t))ds \; dW(t) \\
&= \int_0^{1-\varepsilon} F_i(t)dW(t).
\end{aligned}
$$

Thus

$$
\begin{aligned}
E[(\operatorname{vec} V_i)(\operatorname{vec} V_i)'] &= \int_0^{1-\varepsilon} F_i(t) F_i(t)' dt \\
E[(\operatorname{vec} V_i)(\operatorname{vec} Z)'] &= \int_0^{1-\varepsilon} F_i(t)(S_z \otimes \Psi(t))' dt
\end{aligned}
$$

since $\operatorname{vec}(Z - Z_\varepsilon) = \int_{1-\varepsilon}^1 (I_r \otimes \Psi(t)) S_z dW(t)$ is independent of $\operatorname{vec} V_i$. A direct calculation now shows that $\int_0^{1-\varepsilon} F_a(t) F_a(t)' dt = A_a(I_n \otimes \Sigma_q^\varepsilon) A_a'$, $\int_0^{1-\varepsilon} F_a(t)(S_z \otimes \Psi(t))' dt = A_a(S_z' \otimes \Sigma_q^\varepsilon)$, $\int_0^{1-\varepsilon} F_b(t) F_b(t)' dt = A_b \operatorname{diag}(I_r \otimes \Sigma_q^\varepsilon, I_{kq}) A_b'$ and $\int_0^{1-\varepsilon} F_b(t)(S_z \otimes \Psi(t))' dt = \rho A_b(S_z' \otimes \Sigma_q^\varepsilon)$, so that from the definitions of A_i, $E[(\operatorname{vec} V_i)(\operatorname{vec} V_i)'] = \Sigma_{22}$ and $E[(\operatorname{vec} V_i)(\operatorname{vec} Z)'] = \Sigma_{21}$, as required.

It thus remains to show that the processes $\int_0^s H_i(s,t) dW(t)$, $i \in \{a, b\}$, admit a cadlag version.

Recall that $||A||$ is the Frobenius norm of the real matrix A, $||A|| = \sqrt{\operatorname{tr} A'A}$, which is submultiplicative. If $v \sim \mathcal{N}(0, \Sigma)$, then $E[||v||^4] = E[(v'v)^2] = 2\operatorname{tr}(\Sigma^2) + (\operatorname{tr}\Sigma)^2 \le 3(\operatorname{tr}\Sigma)^2$, so that with $\int_s^t H_i(u,\lambda) dW(\lambda) \sim \mathcal{N}(0, \int_s^t H_i(u,\lambda) H_i(u,\lambda)' d\lambda)$, we find

$$
\begin{aligned}
E[|| \int_s^t H_i(u,\lambda) dW(\lambda)||^4] &\le 3(\operatorname{tr} \int_s^t H_i(u,\lambda) H_i(u,\lambda)' d\lambda)^2 \\
&\le 3(\int_s^t ||H_i(u,\lambda)||^2 d\lambda)^2.
\end{aligned}
$$

Thus, for $0 \le s < t \le 1$, we have with $\psi(s) = d\Psi(s)/ds$

$$
\begin{aligned}
& E[|| \int_0^t H_i(t,\lambda) dW(\lambda) - \int_0^s H_i(s,\lambda) dW(\lambda)||^4] \\
={}& E[|| \int_0^s (H_i(t,\lambda) - H_i(s,\lambda)) dW(\lambda) + \int_s^t H_i(t,\lambda) dW(\lambda)||^4] \\
\le{}& 3[\int_0^s ||H_i(t,\lambda) - H_i(s,\lambda)||^2 d\lambda + \int_s^t ||H_i(t,\lambda)||^2 d\lambda]^2 \\
\le{}& 3[k^2(\sup_{0 \le \lambda \le 1-\varepsilon} ||J(\lambda)^{-1}||^2 ||F_i(\lambda)||^2)(||\Psi(s) - \Psi(t)||^2 + (t-s)\sup_{0 \le \lambda \le 1} ||\Psi(\lambda)||^2)]^2 \\
\le{}& 3k^4(\sup_{0 \le \lambda \le 1-\varepsilon} ||J(\lambda)^{-1}||^4 ||F_i(\lambda)||^4)(\sup_{0 \le \lambda \le 1} ||\psi(\lambda)||^2 + \sup_{0 \le \lambda \le 1} ||\Psi(\lambda)||^2)^2 (t-s)^2
\end{aligned}
$$

where the last inequality follows from $\Psi(t) - \Psi(s) = (t-s) \int_0^1 \psi(s + \lambda(t-s)) d\lambda$, so that by Kolmogorov's continuity theorem (p. 14 of Oksendal (2000)), there exist continuous (and thus cadlag) versions of the stochastic processes $\int_0^s H_i(s,t) dW(t)$, $i \in \{a, b\}$.

37

A.4 Parameterization of $\Sigma_{(Y,X)}$ under H_0 in the Diagonal G-model, and in the G-model when $r > k$

G-model with $r > k$: Because of invariance, it is without loss of generality to assume that the first $r - k$ rows of R are equal to zero, so that the first $r - k$ columns of Z are independent of V. The joint distribution of V and the last k rows of Z are then just as in the model with $r = k$, so that Lemma 1 implies that in the G-model with $r > k$, $\Sigma_{(Y,X)}$ is of the form $\Sigma_{(Y,X)} = \mathrm{diag}(I_{r-k} \otimes I_q, \Sigma_k^*)$ under the null hypothesis, where Σ_k^* is any positive definite $k^2 q \times k^2 q$ matrix with upper left $kq \times kq$ block equal to the identity matrix. The nuisance parameter θ is thus of dimension $k^2 q^2 + kq(kq+1)/2$

Diagonal G-model: Let Z_V and ζ be $q \times k$ random matrices with $\mathrm{vec}(Z, Z_V, \zeta) \sim \mathcal{N}(0, \Sigma_{(Z,Z_V,\zeta)})$, where

$$\Sigma_{(Z,Z_V,\zeta)} = \mathrm{diag}\left(\begin{bmatrix} I_r & R \\ R' & I_k \end{bmatrix}, I_k \right) \otimes I_q.$$

A construction as in the proof of Lemma 1 implies that the j'th column of V can be chosen as an arbitrary linear combination of the j'th column of Z_V and the j'th column of ζ, $j = 1, \cdots, k$ (subject to the constraint that the resulting matrix is positive definite). Thus, $\Sigma_{(Y,X)}$ may be parametrized as $\Sigma_{(Y,X)} = A_{(Z,Z_V,\zeta)} \Sigma_{(Z,Z_V,\zeta)} A'_{(Z,Z_V,\zeta)}$, where

$$A_{(Z,Z_V,\zeta)} = \begin{pmatrix} I_{rq} & 0 & 0 \\ 0 & \mathrm{diag}(A_{V,1}, A_{V,2}, \cdots, A_{V,k}) & \mathrm{diag}(L_{\zeta,1}, L_{\zeta,2}, \cdots, L_{\zeta,k}) \end{pmatrix},$$

$A_{V,j}$ are arbitrary $q \times q$ matrices and $L_{\zeta,j}$ are arbitrary lower diagonal $q \times q$ matrices. Including R, θ is thus of dimension $rk + kq^2 + kq(q+1)/2$.

A.5 Algorithm for Approximating the Least Upper Power Bound and Optimal Test

A computationally more convenient variation of the size adjustment idea described in the main text is as follows: Starting from the level α test $\varphi_\Lambda = \mathbf{1}[\mathrm{LR}_\Lambda > \mathrm{cv}_\Lambda]$ of H_Λ against H_1, for some small $\varepsilon > 0$, let $\mathrm{cv}_\Lambda^\varepsilon$ be an adjusted critical value so that the resulting test $\varphi_\Lambda^\varepsilon = \mathbf{1}[\mathrm{LR}_\Lambda > \mathrm{cv}_\Lambda^\varepsilon]$ (with $\mathrm{cv}_\Lambda^\varepsilon > \mathrm{cv}_\Lambda$) has only slightly lower power than φ_Λ, i.e. $\int \varphi_\Lambda^\varepsilon h d\mu = \beta_\Lambda - \varepsilon$. Now *if* $\varphi_\Lambda^\varepsilon$ is of level α under H_0, i.e. $\sup_{\theta \in \Theta} \int \varphi_\Lambda^\varepsilon f_\theta d\mu < \alpha$, then we have a level

α test of H_0 against H_1 with power that is only ε below β_Λ, and the least upper bound is again sandwiched between β_Λ and $\beta_\Lambda - \varepsilon$. The advantage of this method over the direct size adjustment discussed in the text is that the size adjustment is costly to compute, while Λ can often be quickly dismissed by checking its size control for a small number of values of θ under H_0.

Now consider discrete distributions for Λ: Let $\Theta^N = \{\theta_1, \cdots, \theta_N\} \subset \Theta$ for some $N > 1$ and consider the restricted null hypothesis H_N :The density of U is f_θ, $\theta \in \Theta^N$. In this restricted problem, the least favorable distribution is fully described by the point masses p_i^* on θ_i, $i = 1, \cdots, N$, where $\sum_{i=1}^N p_i^* = 1$. The resulting test φ_N^* is thus of the form $\varphi_N^* = \mathbf{1}[\sum_{i=1}^N p_i^* f_{\theta_i}/h < 1/\operatorname{cv}_N]$. Note that by construction, the test φ_N^* is of level α on $\Theta^N \subset \Theta$. The central idea of the algorithm is to identify a (hopefully not too large) set of points Θ^N so that corresponding adjusted test $\varphi_N^{*\varepsilon}$ is of level α on the whole set Θ.

Introduce the notation $\varphi(\bar{\theta}, \bar{p}, \operatorname{cv})(u)$ for the test $\varphi = \mathbf{1}[\sum_{i=1}^N p_i f_{\theta_i}/h < 1/\operatorname{cv}]$ evaluated at u, with $\bar{\theta} = (\theta_1, \cdots, \theta_N)'$ and $\bar{p} = (p_1, \cdots, p_N)'$, and $\sum_{i=1}^N p_i = 1$ (but $\bar{p}$ is not necessarily the least favorable distribution on Θ^N). The rejection probability of φ under the alternative is $\Pi_1(\bar{\theta}, \bar{p}, \operatorname{cv}) = \int \varphi(\bar{\theta}, \bar{p}, \operatorname{cv})(u)h(u)d\mu(u)$, and it is $\Pi_0(\bar{\theta}, \bar{p}, \operatorname{cv}; \theta)(u) = \int \varphi(\bar{\theta}, \bar{p}, \operatorname{cv})f_\theta(u)d\mu(u)$ under the null hypothesis with $\theta \in \Theta$. We numerically approximate $\Pi_0(\bar{\theta}, \bar{p}, \operatorname{cv}; \theta)$ by

$$\hat{\Pi}_0(\bar{\theta}, \bar{p}, \operatorname{cv}; \theta) = \frac{1}{m} \sum_{j=1}^m \Upsilon_L \left(\frac{\sum_{i=1}^N p_i f_{\theta_i}(u_j)}{h(u_j)}, \frac{1}{\operatorname{cv}} \right) \tag{30}$$

where for some real $L > 0$, $\Upsilon_L : \mathbb{R}^2 \mapsto \mathbb{R}$ is defined as $\Upsilon_L(x, y) = y^L/(y^L + x^L)$. The pseudo random variables u_j, $j = 1, \cdots, m$ have density f_θ and are obtained by suitably transforming some underlying pseudo random variables ξ_j, $u_j = g_\theta(\xi_j)$, $j = 1, \cdots, m$. The variables ξ_j are drawn only once in the evaluation of $\hat{\Pi}_0$ at different arguments (so the transformation g_θ depends on θ). Define $\hat{\Pi}_1(\bar{\theta}, \bar{p}, \operatorname{cv})$ analogously, with u_j given by $g_h(\xi_j)$. Note that as $L \to \infty$, $\Upsilon_L(x, y) \to \mathbf{1}[x < y] + \frac{1}{2}\mathbf{1}[x = y]$, so that for L large, $\hat{\Pi}_0(\bar{\theta}, \bar{p}, \operatorname{cv}; \theta)$ approximates the standard Monte Carlo integration for the rejection probability of $\varphi(\bar{\theta}, \bar{p}, \operatorname{cv})$. The advantage of choosing $L < \infty$ is that $\hat{\Pi}_0(\bar{\theta}, \bar{p}, \operatorname{cv}; \theta)$ and $\hat{\Pi}_1(\bar{\theta}, \bar{p}, \operatorname{cv})$ become smooth and differentiable functions of their arguments, which greatly simplifies numerical optimizations. The computations in this paper were performed with $m = 25,000$ and $L = 25$.

The algorithm consists of three subroutines SR1, SR2 and SR3.

SR1 The routine takes $\bar{\theta} = (\theta_1, \cdots, \theta_N)$ as given, and returns an estimate of the least fa-

vorable distribution on Θ^N, as described by $\bar{p} = (p_1, \cdots, p_N)$. By Theorem 3.8.1 of Lehmann and Romano (2005), the least favorable distribution $\bar{p}^* = (\bar{p}_1^*, \cdots, \bar{p}_N^*)'$ has the two properties (i) $\int \varphi(\bar{\theta}, \bar{p}^*, \mathrm{cv}) f_{\theta_l} d\mu \leq \alpha$ for $l = 1, \cdots, N$; and (ii) $\int \varphi(\bar{\theta}, \bar{p}^*, \mathrm{cv}) f_{\theta_l} d\mu < \alpha$ only if $p_l = 0$. This motivates the joint determination of $\bar{p}$ and cv as numerical solutions to

$$\hat{\Pi}_0(\bar{\theta}, \bar{p}, \mathrm{cv}; \theta_l) \leq \alpha \quad \text{and} \quad p_l(\hat{\Pi}_0(\bar{\theta}, \bar{p}, \mathrm{cv}; \theta_l) - \alpha) = 0 \quad \text{for } l = 1, \cdots, N. \tag{31}$$

Specifically, we determine appropriate $\bar{p}$ and cv by minimizing the objective function

$$\sum_{l=1}^{N} (a_0 p_l + \exp[a_1(\hat{\Pi}_0(\bar{\theta}, \bar{p}, \mathrm{cv}; \theta_l) - \alpha)])(\hat{\Pi}_0(\bar{\theta}, \bar{p}, \mathrm{cv}; \theta_l) - \alpha)^2 \tag{32}$$

where $a_0 = 100$ and $a_1 = 2000$. As a function of $\bar{p}$ and cv, (32) is continuous and with known first derivative, so that a standard quasi-Newton optimizer can be employed. Also, the $N^2 m$ numbers $f_{\theta_i}(g_{\theta_l}(\xi_j))/h(g_{\theta_l}(\xi_j))$ for $i = 1, \cdots, N$, $l = 1, \cdots, N$ and $j = 1, \cdots, m$ can be computed and stored once to speed up the the evaluation of $\hat{\Pi}_0(\bar{\theta}, \bar{p}, \mathrm{cv}; \theta_i)$ and its partial derivatives.

SR2 The routine takes $(\bar{\theta}, \bar{p})$ as inputs and returns vectors $(\bar{\theta}_1, \bar{p}_1)$ of length $N_1 \leq N$ by eliminating pairs of values (θ_j, p_j) with p_j approximately equal to zero.

SR3 The routine takes $(\bar{\theta}, \bar{p})$ as given and either identifies $(\bar{\theta}, \bar{p})$ as yielding a sufficiently precise approximation to the least favorable distribution, or it returns a parameter value $\hat{\theta} \in \Theta$ that needs to be included in the set of points Θ^N. Specifically, the routine consists of three steps:

(a) Solve for the real number cv_Λ that satisfies $\hat{\Pi}_0(\bar{\theta}, \bar{p}, \mathrm{cv}_\Lambda; \theta_l) \leq \alpha$ for all $l = 1, \cdots, N$, so that the test $\varphi(\bar{\theta}, \bar{p}, \mathrm{cv}_\Lambda)$ is the (approximate) level α likelihood ratio test of $H_\Lambda : U$ has density $\sum_{l=1}^{N} p_l f_{\theta_l}$ against H_1.

(b) Compute $\beta_\Lambda = \hat{\Pi}_1(\bar{\theta}, \bar{p}, \mathrm{cv}_\Lambda)$, and numerically solve for $\mathrm{cv}_\Lambda^\varepsilon \geq \mathrm{cv}_\Lambda$ such that $\hat{\Pi}_1(\bar{\theta}, \bar{p}, \mathrm{cv}_\Lambda) - \hat{\Pi}_1(\bar{\theta}, \bar{p}, \mathrm{cv}_\Lambda^\varepsilon) = \varepsilon$. By Lemma 2, β_Λ is (an estimate of) a power bound on level α tests of H_0. As described above, the size adjustment as a function of the power implies that *if* $\varphi(\bar{\theta}, \bar{p}, \mathrm{cv}_\Lambda^\varepsilon)$ is of level α under H_0, then we have found a test whose power is within ε of the bound. The computations in this paper use $\varepsilon = 2.5\%$.

(c) Check on a grid of values $\Theta_G \subset \Theta$ whether $\varphi(\bar{\theta}, \bar{p}, \mathrm{cv}_\Lambda^\varepsilon)$ controls size, i.e. evaluate $\hat{\Pi}_0(\bar{\theta}, \bar{p}, \mathrm{cv}; \theta_j)$ for all $\theta \in \Theta_G$ in the finite set Θ_G. If $\hat{\Pi}_0(\bar{\theta}, \bar{p}, \mathrm{cv}; \theta_j) > \alpha$ for some j, return $\hat{\theta} = \theta_j$. Otherwise, preliminarily accept β_Λ as the approximate least upper bound, and $\varphi(\bar{\theta}, \bar{p}, \mathrm{cv}_\Lambda^\varepsilon)$ as an approximately efficient test. As a practical matter, it makes sense to return $\hat{\theta}_j = \theta_j$ even if $\hat{\Pi}_0(\bar{\theta}, \bar{p}, \mathrm{cv}; \theta_j)$ is below, but very close to α. We use a threshold of 4.8% for $\alpha = 5\%$.

Overall the algorithm iterates between the subroutines as follows:

1. Initialize $\bar{\theta}$ with $N = 25$ values of θ_j that are spread out over the grid Θ_G, and call SR1.

2. Call SR2 to obtain a new N and $(\bar{\theta}, \bar{p})$ pair.

3. While SR3 returns $\hat{\theta}$:

   Add $\hat{\theta}$ to $\bar{\theta}$ (so that N is increased by one) and call SR1.

4. Call SR2 to obtain a new N and $(\bar{\theta}, \bar{p})$ pair. Repeat Step 3.

5. Perform a final check on whether $\varphi(\bar{\theta}, \bar{p}, \mathrm{cv}_\Lambda^\varepsilon)$ is a level α test by evaluating its rejection probability over a fine grid of values for θ_j, using a different set of draws of pseudo-random variables ξ_j in (30). For the correlation R, we use the grid $R = -0.99, -0.96, \cdots, 0.99$ in the $I(1)$ model, and in the local-to-unity model, a square grid of the same values of R, and $C = -3, -2.5, \cdots, -.5, 0, e^{-1}, e^{-0.8}, \cdots, e^{6.8}, e^7$.

The advantage of the thinning operation in SR2 is that it accelerates the computation of the test statistic, and it facilitates the numerical minimization of (32). We do not call SR2 after each call of SR1, though, because doing so can result in cycles, so that Step 3 above could potentially result in an infinite loop. Step 4 "cleans" the feasible test and is skipped when only an estimator of the least upper bound is required.

References

ANDREWS, D. W. K., M. J. MOREIRA, AND J. H. STOCK (2007): "Efficient Two-Sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments," *Working paper, Harvard University.*

BIERENS, H. J. (1997): "Nonparametric Cointegration Analysis," *Journal of Econometrics*, 77, 379–404.

BOBKOSKI, M. J. (1983): "Hypothesis Testing in Nonstationary Time Series," *unpublished Ph.D. thesis, Department of Statistics, University of Wisconsin.*

CAMPBELL, J. Y., AND M. YOGO (2006): "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81, 27–60.

CANER, M., AND L. KILIAN (2001): "Size Distortions of Tests of the Null Hypothesis of Stationarity: Evidence and Implications for the PPP Debate," *Journal of International Money and Finance*, 20, 639–657.

CANJELS, E. (1997): "Essays on Time Series Econometrics and Sharecropping in Agriculture," *Ph.D. Thesis, Northwestern University.*

CAVANAGH, C. L. (1985): "Roots Local To Unity," *Working Paper, Harvard University.*

CAVANAGH, C. L., G. ELLIOTT, AND J. H. STOCK (1995): "Inference in Models with Nearly Integrated Regressors," *Econometric Theory*, 11, 1131–1147.

CHAN, N. H., AND C. Z. WEI (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15, 1050–1063.

ELIASZ, P., J. H. STOCK, AND M. W. WATSON (2004): "Optimal Tests for Reduced Rank Time Variation in Regression Coefficients and for Level Variation in the Multivariate Local Level Model," *Working paper, Princeton University.*

ELLIOTT, G. (1998): "The Robustness of Cointegration Methods When Regressors Almost Have Unit Roots," *Econometrica*, 66, 149–158.

ELLIOTT, G., AND J. H. STOCK (1994): "Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown," *Econometric Theory*, 10, 672–700.

GRANGER, C. W. J. (1993): "What Are We Learning About the Long-Run?," *The Economic Journal*, 103, 307–317.

HARVEY, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

JANSSON, M. (2004): "Stationarity Testing with Covariates," *Econometric Theory*, 20, 56–94.

JANSSON, M., AND M. J. MOREIRA (2006): "Optimal Inference in Regression Models with Nearly Integrated Regressors," *Econometrica*, 74, 681–714.

JEGANATHAN, P. (1997): "On Asymptotic Inference in Linear Cointegrated Time Series Systems," *Econometric Theory*, 13, 692–745.

——— (1999): "On Asymptotic Inference in Cointegrated Time Series with Fractionally Integrated Errors," *Econometric Theory*, 15, 583–621.

JOHANSEN, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

KEMPTHORNE, P. J. (1987): "Numerical Specification of Discrete Least Favorable Distributions," *SIAM Journal on Scientific and Statistical Computing*, 8, 171–184.

KIM, C. S., AND P. C. B. PHILLIPS (2000): "Fully Modified Estimation of Fractional Cointegration Models," *Working Paper, Yale University*.

KWIATKOWSKI, D., P. C. B. PHILLIPS, P. SCHMIDT, AND Y. SHIN (1992): "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics*, 54, 159–178.

LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses*. Springer, New York.

MAGNUS, J. R., AND H. NEUDECKER (1988.): *Matrix Differential Calculus*. Wiley, New York.

MÜLLER, U. K. (2005): "Size and Power of Tests of Stationarity in Highly Autocorrelated Time Series," *Journal of Econometrics*, 128, 195–213.

——— (2007a): "An Alternative Sense of Asymptotic Efficiency," *Working paper, Princeton University.*

——— (2007b): "A Theory of Robust Long-Run Variance Estimation," *Journal of Econometrics,* 141, 1331–1352.

MÜLLER, U. K., AND M. W. WATSON (2007): "Testing Models of Low-Frequency Variability," *forthcoming in Econometrica.*

NABEYA, S. (1951): "Absolute Moments in 2-Dimensional Normal Distribution," *Annals of the Institute of Statistical Mathematics,* 3, 2–6.

NELSON, W. (1966): "Minimax Solution of Statistical Decision Problems by Iteration," *The Annals of Mathematical Statistics,* 37, 1643–1657.

NYBLOM, J. (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association,* 84, 223–230.

NYBLOM, J., AND A. HARVEY (2000): "Tests of Common Stochastic Trends," *Econometric Theory,* 16, 176–199.

NYBLOM, J., AND T. MÄKELÄINEN (1983): "Comparisons of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model," *Journal of the American Statistical Association,* 78, 856–864.

OGAKI, M. (1988): "Learning About Preferences from Time Trends," *Unpublished Ph.D. Dissertation, University of Chicago.*

OGAKI, M., AND J. Y. PARK (1998): "A Cointegration Approach to Estimating Preference Parameters," *Journal of Econometrics,* 82, 107–134.

OKSENDAL, B. (2000): *Stochastic Differential Equations: An Introduction with Applications.* Springer, Berlin, 5 edn.

PARK, J. Y. (1992): "Canonical Cointegrating Regressions," *Econometrica,* 60, 119–143.

PHILLIPS, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika,* 74, 535–547.

———— (1998): "New Tools for Understanding Spurious Regression," *Econometrica*, 66, 1299–1325.

PHILLIPS, P. C. B., AND B. E. HANSEN (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies*, 57, 99–125.

ROBINSON, P. M., AND J. HUALDE (2003): "Cointegration in Fractional Systems with Unknown Integration Orders," *Econometrica*, 71, 1727–1766.

ROBINSON, P. M., AND D. MARINUCCI (2003): "Semiparametric Frequency Domain Analysis of Fractional Cointegration," in *Time Series with Long Memory*, ed. by P. M. Robinson, pp. 334–373. Oxford University Press, Oxford.

SAIKKONEN, P. (1991): "Estimation and Testing of Cointegrating Regressions," *Econometric Theory*, 7, 1–27.

SAIKKONEN, P., AND R. LUUKONEN (1993): "Point Optimal Tests for Testing the Order of Differencing in ARIMA Models," *Econometric Theory*, 9, 343–362.

SRIANANTHAKUMAR, S., AND M. L. KING (2006): "A New Approximate Point Optimal Test of a Composite Null Hypothesis," *Journal of Econometrics*, 130, 101–122.

STOCK, J. H., AND M. W. WATSON (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783–820.

———— (1996): "Confidence Sets in Regressions with Highly Serially Correlated Regressors," *Working Paper, Harvard University*.

VELASCO, C. (2003): "Gaussian Semiparametric Estimation of Fractional Cointegration," *Journal of Time Series Analysis*, 24, 345–378.

WRIGHT, J. H. (2000): "Confidence Sets for Cointegrating Coefficients Based on Stationarity Tests," *Journal of Business and Economic Statistics*, 18, 211–222.

Table 1: 1%, 5%, and 10% Critical Values for the *JW* Statistic

q	*r*				
	1	**2**	**3**	**4**	**5**
6	5.25 3.62 3.08	6.76 5.16 4.39	7.25 6.09 5.43	7.14 6.46 6.02	6.54 6.33 6.16
7	4.33 3.08 2.68	5.52 4.20 3.63	6.05 4.95 4.37	6.15 5.35 4.92	5.89 5.48 5.20
8	3.68 2.73 2.39	4.65 3.54 3.08	5.17 4.16 3.68	5.29 4.55 4.12	5.26 4.73 4.42
9	3.21 2.46 2.18	4.02 3.09 2.73	4.46 3.58 3.19	4.63 3.93 3.56	4.66 4.12 3.83
10	2.86 2.25 2.02	3.56 2.79 2.48	3.94 3.17 2.84	4.10 3.47 3.15	4.18 3.66 3.38
11	2.62 2.10 1.90	3.16 2.54 2.29	3.53 2.87 2.59	3.71 3.12 2.84	3.78 3.30 3.03
12	2.46 1.98 1.81	2.89 2.35 2.13	3.18 2.64 2.39	3.38 2.84 2.60	3.48 3.02 2.78
13	2.29 1.88 1.73	2.68 2.21 2.01	2.92 2.44 2.23	3.13 2.63 2.42	3.20 2.77 2.57
14	2.16 1.80 1.67	2.50 2.09 1.92	2.74 2.31 2.11	2.91 2.47 2.27	2.97 2.59 2.40
15	2.07 1.74 1.61	2.36 1.99 1.84	2.56 2.18 2.01	2.69 2.32 2.15	2.80 2.44 2.27
16	1.97 1.67 1.56	2.24 1.91 1.77	2.44 2.08 1.92	2.55 2.21 2.05	2.64 2.30 2.15
17	1.89 1.62 1.52	2.15 1.84 1.71	2.32 1.99 1.85	2.43 2.11 1.96	2.50 2.20 2.05
18	1.82 1.58 1.49	2.07 1.78 1.66	2.21 1.92 1.79	2.32 2.02 1.89	2.39 2.10 1.98

Note: The table shows asymptotic critical for the $JW(b)$ statistic computed using $b = 10/\sqrt{r}$, where $JW(b) = \det(Y'Y)/\det(Y'(I + b^2D)^{-1}Y)$, with $D = \mathrm{diag}(d_1,...,d_k)$ and $d_i = (i\pi)^{-2}$.

Table 2: Asymptotic power bounds for best invariant tests under different restrictions on the common trend process under the null hypothesis for $q = 12$

A. $r = k = 1$ (KLIC and LUB)

	KLIC Minimization				Least Upper Bound	
	Restriction on Trend Process Under the Null Hypothesis					
	None	*Stationary*	*Local-to-Unity*	*I(1)*	*Local-to-Unity*	*I(1)*
	(i) $|B| = 7$, $JW(|B|)$ power = 0.36, $JW(10)$ power = 0.36					
$|R| = 0.0$	0.36	0.41	0.50	0.50	0.50	0.50
$|R| = 0.5$	0.36	0.54 0.40	0.66 0.59	0.65 0.65	0.66 0.58	0.65 0.65
$|R| = 0.9$	0.36	0.89 0.44	0.96 0.67	0.95 0.95	0.93 0.65	0.94 0.94
	(ii) $|B| = 14$, $JW(|B|)$ power = 0.64, $JW(10)$ power = 0.63					
$|R| = 0.0$	0.64	0.69	0.81	0.82	0.78	0.81
$|R| = 0.5$	0.64	0.81 0.67	0.93 0.80	0.92 0.92	0.88 0.78	0.90 0.90
$|R| = 0.9$	0.64	0.99 0.72	1.00 0.87	1.00 1.00	1.00 0.82	1.00 1.00

B: $r = 1$ and $k \geq 2$ (KLIC minimization for $k = 2$)

	Restriction on Trend Process Under the Null Hypothesis				
	None	*Diagonal G(s,t)*	*Diagonal and Stationary G(s,t)*	*Local-to-Unity*	*I(1)*
	$\|B\| = 7$, $JW(\|B\|)$ power = 0.36, $JW(10)$ power = 0.36				
$\|R\| = 0.0$	0.36	0.36	0.39	0.42	0.42
$\|R\| = 0.5$	0.36	0.36 0.37 0.36	0.47 0.42 0.38	0.54 0.48 0.51	0.54 0.47 0.54
$\|R\| = 0.9$	0.36	0.36 0.43 0.36	0.82 0.59 0.43	0.92 0.69 0.64	0.92 0.86 0.92
	$\|B\| = 14$, $JW(\|B\|)$ power = 0.64, $JW(10)$ power = 0.63				
$\|R\| = 0.0$	0.64	0.64	0.67	0.71	0.71
$\|R\| = 0.5$	0.64	0.65 0.64 0.64	0.76 0.71 0.66	0.83 0.77 0.76	0.82 0.77 0.83
$\|R\| = 0.9$	0.64	0.66 0.74 0.66	0.97 0.84 0.71	1.00 0.85 0.85	0.99 0.98 0.99

C: $r = 2$ and $k = 1$ (KLIC minimization)

	Restriction on Trend Process Under the Null Hypothesis					*Y-only Power Envelope*
	None	*G(s,t)*	*Stationary G(s,t)*	*Local-to-Unity*	*I(1)*	
	$\|B\| = 10$, $JW(10/\sqrt{2})$ power = 0.39					
$\|R\| = 0.0$	0.43	0.43	0.49	0.54	0.55	0.41
$\|R\| = 0.5$	0.46 0.49 0.47	0.46 0.54 0.47	0.61 0.59 0.50	0.72 0.59 0.65	0.69 0.59 0.70	0.41 0.44 0.41
$\|R\| = 0.9$	0.50 0.69 0.50	0.50 0.74 0.50	0.94 0.74 0.57	0.98 0.74 0.76	0.97 0.74 0.97	0.41 0.55 0.41
	$\|B\| = 20$, $JW(10/\sqrt{2})$ power = 0.58					
$\|R\| = 0.0$	0.68	0.68	0.71	0.78	0.83	0.64
$\|R\| = 0.5$	0.72 0.74 0.72	0.72 0.76 0.72	0.81 0.77 0.74	0.91 0.85 0.81	0.91 0.87 0.91	0.64 0.68 0.64
$\|R\| = 0.9$	0.79 0.94 0.79	0.79 0.97 0.79	0.98 .097 0.83	1.00 0.97 0.90	1.00 0.97 1.00	0.64 0.86 0.64

Notes: Entries are power bounds for 5% level tests against $I(1)$ alternative described by B and R, where tests are restricted to control size for trend process listed in the column headings. Side-by-side entries correspond to $R \cdot B < 0$ and $R \cdot B > 0$ in Panel A, and to $tr(R'B)/(\|R\| \cdot \|B\|) = \{-1, 0, 1\}$ in Panels B and C. KLIC minimization refers to bounds constructed using optimal tests for simple null hypotheses satisfying the KLIC minimization problem from section 5.4.2. In Panel A, the bounds under the heading Least Upper Bound were computed using the approximation discussed in Section 5.4.1.

Table 3: Asymptotic Power of Feasible Weighted Average Power Maximizing Tests for $r = k = 1$ and $q = 12$

	Restriction on Trend Process Under the Null Hypothesis			
	I(1)		Local-to-Unity	
	Feasible Test	Envelope	Feasible Test	Envelope
	$\|B\| = 7$			
$\|R\| = 0.0$	0.47	0.50	0.46	0.50
$\|R\| = 0.5$	0.53 0.53	0.65 0.65	0.55 0.48	0.66 0.58
$\|R\| = 0.9$	0.87 0.87	0.94 0.94	0.88 0.61	0.93 0.65
	$\|B\| = 14$			
$\|R\| = 0.0$	0.74	0.81	0.70	0.78
$\|R\| = 0.5$	0.79 0.79	0.90 0.91	0.79 0.71	0.88 0.78
$\|R\| = 0.9$	0.97 0.97	1.00 1.00	0.97 0.73	1.00 0.82

Notes: Entries below Feasible Test tabulate power of two feasible 5% level tests that control size in the I(1) and local-to-unity model, respectively, against the I(1) alternative described by B and R. By construction, these two tests approximately maximize equally weighted average power against the I(1) trend model with parameters $B = 7$ and $R = -0.95, -0.85, \cdots, 0.85, 0.95$. Entries below Envelope tabulate the approximate power envelopes from Table 2 for comparison. Side-by-side entries correspond to $R \cdot B < 0$ and $R \cdot B > 0$.

Figure 1: Asymptotic power of $JW(b)$ test and power envelope for Y–only tests for $I(1)$ alternative, $R = 0$ and $q = 12$